Lawrence E. Murr

Handbook of Materials Structures, Properties, Processing and Performance

Springer Reference

Handbook of Materials Structures, Properties, Processing and Performance

Lawrence E. Murr

Handbook of Materials Structures, Properties, Processing and Performance

With 784 Figures and 43 Tables

Springer Reference

Lawrence E. Murr
Metallurgical and Materials Engineering
The University of Texas at El Paso
El Paso, TX, USA

ISBN 978-3-319-01814-0 ISBN 978-3-319-01815-7 (eBook)
ISBN 978-3-319-01816-4 (print and electronic bundle)
DOI 10.1007/978-3-319-01815-7
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014954230

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Pat, my wife and soul mate
of more than 55 years at this writing,
and Brayden Lawrence Blanks,
our great grandson.

Editorial Advisory Board

Preface

This handbook is an attempt at innovation as it applies to contemporary and evolving materials science and engineering principles and their many applications. It is intended to be interdisciplinary and multidisciplinary as these apply generally to the realm of science and technology encumbered by materials science and engineering in its broadest sense. This work could have been titled many other broader or specific terms, although handbook itself invokes many interrelated definitions which in fact provide a broader context as it applies to materials structures, properties, processing, and performance. For example, consider the following *handbook* definitions:

- A type of reference work or guide book.
- A treatise on a special subject.
- A concise manual or reference book providing specific information or instruction about a subject.
- Comprehensive and detailed work on a particular topic for practitioners, Structured for quick reference and often used as a supplement to a text book.

This handbook ambitiously strives to be all of these. In addition, it can not only serve as a supplement to some specific text, but it can also be configured as a text by selecting specific parts and chapters to be configured for an appropriate course at the graduate or advanced level. While many fundamental topics are briefly reviewed in order for the reader or user (instructor or facilitator) to better appreciate some topic areas, especially what might be considered more advanced materials or materials science and engineering concepts, most of this work is not intended to be used as any configuration of principles of, fundamentals of, or introduction to materials science or materials science and engineering. It is intended to provide specifics as well as an overview of many contemporary materials and materials science and engineering–related topic areas and research areas, especially those which might suggest evolving or future areas relating to advanced materials and manufacturing processes and technologies. In this context, this *handbook* is intended as a concise guide to contemporary materials science and engineering for the technical professional, especially those involved in leading-edge research and development as well as technology development and forecasting in the context

of future product directions involving aspects of advanced materials structures, properties, processing, and performance. These are in fact expanded to ultimately include materials architectures and materials design as it is related to materials simulations and more generally computational issues involving materials science and engineering.

Consistent with these strategies and innovations implicit in the broader context of this Handbook, this work will at first appear as a print and online edition (as an eBook on SpringerLink), and in addition an updated/updatable version will be available on www.SpringerReference.com.

This work in some respects represents an effort encompassing a half-century since it incorporates a wide range of the author's research relating to materials fundamentals dating to the mid-1960s. It incorporates many stages of class notes over at least the past two decades used in graduate courses in materials science and engineering, especially as a discipline-leveling course to accommodate graduate students in a wide range of science and engineering disciplines as well as materials and materials science and engineering. In this sense, I am indebted to many hundreds of students and colleagues who have contributed to this work in various ways, many of whom are acknowledged in the text or in figure credits and references from which much of the work has been adapted. This includes hundreds of undergraduate and graduate student researchers who have contributed to publishing many hundreds of papers in the international literature and who have been a part of my extended academic family for nearly half a century.

I am especially indebted to my administrative assistant of more than two decades, Faye Ekberg, who typed much of the initial manuscript, and Pamela Kuchle and Abelardo Castrejon, undergraduate student assistants who scanned the many illustrations and made many corrections and additions to the final manuscripts, especially those tendered by my Editorial Advisory Board: Drs. Marc Meyers, Naresh Thadhani, Shujun Li, Wenjun Zhu, and Sridhar Pappu. These represent colleagues, students, and long-standing friends who go back more than four decades. Finally, I must acknowledge the help of Michael Luby, Publishing Editor, Engineering, Springer Science + Business Media (New York), Lydia Mueller, Springer Publishing Product Development Senior Editor, Major Reference Works (Heidelberg), and Karin Bartsch, Springer Publishing Product Development Associate Editor MRW (Heidelberg), who helped to configure my initial proposal and provided guidance and support in producing the final Handbook in its various forms.

L. E. Murr

Contents

Volume 1

Volume 2

Part I

The Ages of Metals and Materials

A Brief History of Metals

Contents

Abstract

This initial chapter briefly reviews the principal ages of materials and metals: the Stone Age, the Copper Age, the Bronze Age, and the Iron Age, spanning some ten millennia. Incidental to these major materials ages, gold and silver are noted as these were initially utilized in native forms, along with copper and iron, which has been identified throughout antiquity in meteoric forms along with nickel. There were ideally eight metals employed in civilizations in antiquity: copper, gold, silver, iron, tin (which when alloyed with copper heralded the Bronze Age), zinc, lead, and mercury.

Introduction

Our current, wide use of materials – metals, ceramics, plastics/polymers, semiconductors, superconductors, composites, etc. – is often referred to as representing the "Age of Materials." The reality, however, is that the age of materials began with the Stone Age (Fig. 1) dating well before 10,000 B.C. when prehistoric peoples fashioned crude tools, weapons, and other commodities from stone. But stone continues in use today, not as a basis for tools and weapons but as a basis for building materials and other construction applications. The so-called Seven Wonders of the World are all of stone construction, and they exemplify not only the variations in stone as a material but also the selective use of these variations in both the engineering and art of construction. In addition a variety of clays and wood augmented stone materials and remain a significant part of our materials technology.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_1

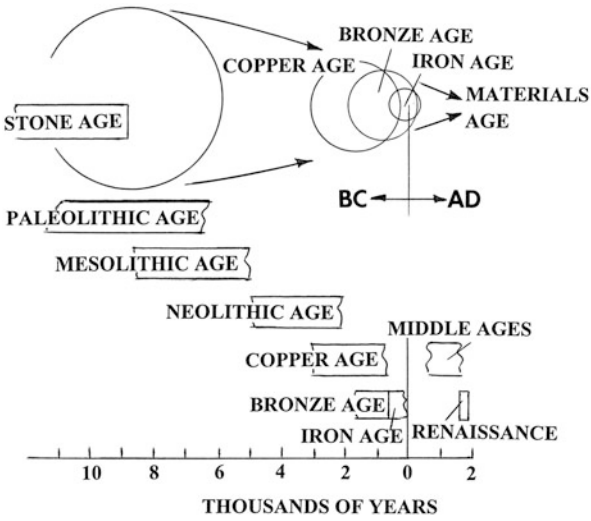


Fig. 1 The Metal Ages, including the Stone Age which dates back to prehistoric times

The earliest metal to be used by people in a significant way was copper, and the world's earliest known man-made copper objects – beads, pins, and awls – were fabricated about 8000 B.C. in the region which is near Turkey and Iran, and often designated as the start of the Copper (or Chalcolithic) Age (Fig. 1). Copper was probably first discovered in its native state in various locations worldwide in prehistoric times, and there is evidence of copper mining in the Balkans circa 5000 B.C. Copper came to prominence in Europe as early as the fourth millennium B.C. when artisans – the first extractive metallurgists – learned to extract it by smelting ore. In this early ore smelting or copper extraction process, copper oxide ore was heated in open and closed rock and clay crucibles or hearths buried in the ground, using wood, sometimes mixed with the ore, as the fuel (Fig. 2). Somehow it was discovered that blowing air through hollow sticks embedded in the fuel-ore mass fanned the fire to create heat intense enough to free the copper as a molten mass. Actually the molten mass consisted of slag which contained impurities from the ore which separated upon cooling and could be chipped away, leaving relatively pure copper metal blocks. These blocks, or ingots, were later reheated, and the liquid copper poured into a cast or mold, or fashioned into tools such as ax blades, and weapons, sharper and more workable than stone. Indeed it is the quality of workability or *malleability* which continued to set copper apart as a unique material over a period of at least 10,000 years, including the current technologies of the world. Even during the period of the earliest uses of native copper, native gold had also been discovered and was being utilized. Somehow it was observed that hammering the two metals caused copper to harden (or *work harden*), whereas gold remained malleable. Those who recognized and utilized these performance qualities were probably the first adaptive metallurgists. This intuition about the physical and mechanical properties of native metals probably dates prior to the Copper Age (circa 6000 B.C.). The demand for copper increased with agricultural

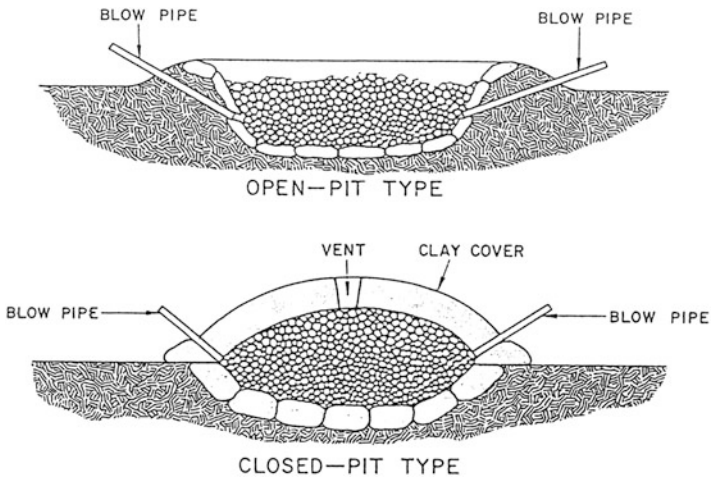


Fig. 2 Early air-blown hearths containing mixtures of metal oxide ore and carbon fuel established the basis for metal extraction from the Copper Age to the Iron Age, spanning a period of more than 5,000 years. In the late Iron Age, this hearth design was modified to become the basis for the blast furnace for steel production. The basis for these processes was that metal oxides plus carbon fuels yielded carbon oxides and free metal. Wood was replaced with charcoal and coal replaced charcoal in the first millennium A.D. Coke became the fuel of choice in more recent times (From "The Making, Shaping, and Treating of Steels," H. E. McGarron, Ed., U. S. Steel Corp., Ninth Edition, 1971)

inventions like the plow and wheeled carts, and traders, prospectors, and early metallurgists exchanged ideas and tool making skills. The addition of tin to copper heralded the transition to the Bronze Age around 3000 B.C.

While metallurgy began with a realization that copper could be hammered into shapes which held sharp edges far better than gold (or even silver which was also discovered as native metal in prehistoric times), the utility of metallurgy in the shaping of civilizations began with the development of systematic processes to extract metal from its ore. Copper smelting – or ore reduction using a hearth – was the basis for metal recovery well into the Iron Age (or the late Bronze Age circa 2000 B.C.).

It was not known in the earliest hearth configurations that enclosed wood not only provided heat sufficient to melt copper (1,981 °F or 1,066 °C) but even more importantly the carbon available in the fuel served to release the metal by first combining with the air blown into the hearth to form carbon monoxide, and then instantaneously, and because of the high temperature, the CO combined with the oxygen in the ore to form carbon dioxide, thereby freeing the copper:

$$2C + O_2 \rightarrow 2CO + 2CuO \rightarrow 2CO_2\uparrow + \underline{2Cu}$$

Sometime after civilizations learned to extract copper from oxide ore, the presence of copper was noticed in mine drainage waters by the Romans, Phoenicians, and Spaniards circa 1000 B.C. where crude iron objects became "coated" with copper when left in these waters. This process, called *cementation*,

is thought to be the origin of alchemy. The extraction of copper by leaching and recovery by deposition onto iron was, however, not to be exploited as an extraction technique until about 1670 A.D. when the first such activity was recorded at the Rio Tinto Mine in Spain. It was also during this period of nearly 3,000 years that copper sulfide ores were roasted to produce an oxide which was reduced to free copper.

During the thousands of years native copper was worked and copper smelting was discovered and developed as a technological process, tin, a white metal, was also somehow smelted. Tin smelting was easier than copper because tin melts at only 232 °C. Somehow, it is believed, a metallurgist in antiquity discovered that tin could be mixed with copper not only to produce a wide range of working options but even alterations in the appearance or luster of copper. The mixing of tin and copper to produce an alloy we call bronze ushered in the Bronze Age (around 3000 B.C.) and created a versatility in the use of copper by lowering the processing temperatures, which made it more castable and more tractable in a wide range of related applications. This phenomenon was expanded with the addition of zinc to tin bronze in the Middle Ages to form castable gun metal (88 % Cu, 10 % Sn, 2 % Zn). In reality, this was the foundation of the "age of alloys" which began a technological phenomena we today call "materials by design." Metal alloys first discovered in the Bronze Age (Fig. 1) influenced the development of an enormous range of metal combinations which attest to essentially all of the engineering achievements of modern times, including atomic energy, space flight, air travel, communications systems and microelectronic devices, and every modern structural and stainless steel building, vessel, or commodity item, including eating utensils.

It is believed that early man found iron meteorite fragments which were shaped by hammering into tools, weapons, and ornaments, because iron is rarely found in the native state, in contrast to copper, gold, or silver. In addition, chemical analysis of archeological specimens often shows 7–15 % nickel, and natural iron-nickel alloys (awaruite, $FeNi_2$, and josephinite, Fe_3Ni_5) are extremely rare and almost exclusively confined to a geological region in northwestern Greenland. Ancient writings from India and China suggest that ferrous metals may have been extracted by the reduction of iron ore in hearth processes similar to copper ore reduction as early as 2000 B.C., and this extraction process was well established over a wide area of the ancient world between about 1400 and 1100 B.C., establishing an era referred to as the Iron Age (Fig. 1). As time went by, variations in hearth or furnace design emerged along with discoveries of variations in the properties of iron brought on by *alloying* due to the absorption of carbon. These included a reduction in the temperature required for the production of molten (*pig*) iron and the production of melted iron used to make castings (*cast iron*) when the carbon content was high (between 3 and 4 wt%). Low carbon metal, in contrast, was relatively soft, ductile, and malleable and could be hammered or hammer-welded at forging temperatures (in an open-hearth furnace) and corresponded to what became generally known as *wrought iron*. Somewhere near 1 wt% carbon (or somewhere between a very low carbon and a very high carbon-containing iron), the iron-carbon alloy produced could be made to exhibit a range of hammer-related characteristics, not the least of which involved extreme hardness when it was variously cooled by

immersion in water or some other liquid from a high temperature. The process of *quenching* was the forerunner of modern steel production. The quenched metal could be reheated for short periods at lower temperature to reduce the hardness and concomitant brittleness in what is now called *tempering*. The metal could also be variously worked in tempering cycles which provided a range of hardness and strength to the final products, forerunners of specialty steels.

During the first 1,500 years of the Iron Age, iron was not melted and correspondingly not castable because the hearth temperatures necessary could not be attained. Around 300 B.C. the Chinese apparently discovered a forced air bellows system that could create temperatures above 1,400 °C which is necessary to melt the iron. This was not achieved in Europe until around 1400 A.D. The Chinese poured molten iron into molds to cast plow shares and hoes. Prior to castable iron, the hammering of the aggregate mixtures of iron and slag (called *blooms*) to separate the iron and slag and produce wrought iron was done by blacksmiths using open hearths for heat.

It was known for centuries that iron ores, mainly oxides, embedded in burning charcoal could be reduced to metallic iron. Oxygen (in air) blown into the bottom of the charcoal bed combines with the carbon to generate heat by forming carbon dioxide which is almost immediately converted to carbon monoxide (Fig. 2). The hot carbon monoxide reacts with oxygen in the iron oxides to reform carbon dioxide which escapes, leaving metallic iron and other impurities in the form of a molten slag which separates from the molten iron:

$$2C + O_2 \rightarrow 2CO + 2\,(FeO)^{1} \rightarrow 2\,CO_2\uparrow + \underline{2Fe}$$

Over the millennia, a great variety of processes were developed which variously mixed or stratified the iron ore and charcoal (or related carbonaceous fuel) and blew air into the combusting mixture. Early hearths were even constructed on hillsides so prevailing winds could provide strong air blasts required for intense combustion by channeling the wind into long tunnels leading to the hearth. Variations of bellows assemblies replaced the natural air blast arrangements, and the furnaces became shaft-like where the raw materials (ore and carbon fuel) could be added continuously from the top of the shaft. The so-called blast furnace, introduced into England around 1500 A.D., was modified about 100 years later to utilize coke[2] as a fuel, and around 1800 A.D. the principle of heating the air before it was blown into the furnace was introduced. The first US iron works was established in Virginia in 1619 on the James River, and the Hammersmith (Saugus) Massachusetts iron works,

[1] FeO does not exist. Common iron oxides include Fe_2O_3, Fe_3O_4, or other stoichiometries which in a mixture might be approximated by FeO.

[2] Coke is to coal as charcoal is to wood, a fuel concentration of nearly pure carbon. Coke is produced by heating coal in the absence of oxygen to drive off volatile components, leaving a mass of strong, porous carbon.

started in 1645, served as the American blast furnace standard up to about 1850. The Bessemer process (1856) and the open-hearth furnace became the basis for making steel worldwide near the end of the 1800s.

When coke became the fuel of choice in blast furnace processes with vertical stacks, it was found that the process could be made more efficient through slag control and modification, which included the slag chemistry, fluidity, and its ability to capture unwanted impurities such as sulfur, which was carried away in slags. Agents to modify the slags became known as fluxes, and limestone ($CaCO_3$) became the principal addition to blast furnaces along with coke and iron ore. Silica in the coke and iron ore and a variety of other oxide minerals in the iron ore produced complex slag chemistries which were studied in great detail to control their corrosiveness and interactions with the ceramic refractories lining the furnaces, ladles, and other retaining vessels, including fireclays, aluminas, and a host of complex oxide systems.

While there are no records of the first uses of metals, it is likely that the occurrence of native or free metals such as gold, silver, and copper over the earth's surface led prehistoric peoples to collecting and variously fashioning them into assorted tools, ornaments, and weapons. With the proliferation and deliberate use of fire, the first to observe the melting of these metals must surely have believed they had encountered a phenomenon akin to magic. It is difficult to imagine how a metal like copper was first extracted from an ore. The observation of a molten metal emerging from a rock mass was probably first observed in some kind of natural occurrence, which was studied over hundreds or thousands of years and finally replicated in a purposeful action by one regarded as a wizard of the time. Descriptions of processes were apparently only sketchy over the few millennia B.C., and details of mining and metallurgy in the European regions only became generally well known in the writings of G. Agricola in De Re Metallica around 1590 A.D. The first comprehensive description of gold metallurgy in all its branches appears in De Re Metallica, and the use of gold for ornamental purposes predates any known human record. The earliest metallurgical working of gold is recorded in Egyptian writings as early as 4000 B.C. De Re Metallica served as the first textbook for miners and metallurgists from its Latin edition of 1556 A.D. until Schlüter's great work on metallurgy in 1738 A.D. De Re Metallica by Georgius Agricola was first written and published in 1556 A.D. Herbert Clark Hoover and Lou Henry Hoover translated this edition, which was published originally in 1912, and republished by Dover Publishing Company in 1950. This work dealt with mining and a wide range of metal recovery processes, which included metals and alloys. Process details were illustrated in nearly 300 illustrations which in themselves represent the art and culture of this medieval time. The footnotes added during the translation provide far-reaching perspectives and fascinating reading (Agricola 1950).

There are usually eight metals identified in antiquity: copper, gold, silver, tin, and iron as discussed, along with zinc, lead, and mercury. These metals were easy to recover from sulfide minerals by smelting as a consequence of the low melting points of zinc and lead. Mercury sulfide (HgS) or cinnabar (often referred to as vermillion) also constituted the first red pigment in antiquity. While Hg can be smelted (or roasted) from this sulfide mineral, mercury and sulfur were often

combined to produce this pigment. Like tin, zinc was later alloyed with copper to make brass (around 1000 A.D.), and lead was cast into numerous forms because of its low melting point in different civilizations in antiquity. Lead is often considered to have been one of the first metals discovered and used after ancients discovered fire. Lead poisoning has also been documented nearly 500 years before the building of the pyramids, and Hippocrates (around 350 B.C.) accurately described the symptoms of lead poisoning.

Artisans of metallurgy have existed for thousands of years and have included the ornament makers and sword makers who have inspired civilizations worldwide. While there may not be detailed scripts for metal production and processing, secrets locked for thousands of years are often revealed by simple surface treatments such as the swabbing of a dilute acid solution onto a metal surface to release patterns frozen in time by a process called etching. For example, the special features of the Damascus swords and daggers have only recently been revealed by metallography. More sophisticated techniques can reveal the source of the metal, the method by which it was produced, and even the age of the metal artifact.

Metallurgists in antiquity were the artisans who could fashion metals into ornaments or other useful products often in a style to meet a specified purpose based on knowledge passed from generation to generation or artisan to artisan. Probably for thousands of years, metallurgists extracted and worked metals. However, when the need to extract large quantities of metals occurred, specialization began to characterize the extraction and production of raw metals in contrast to their hammering and processing to make commodity and structural items. Extraction was often intimately connected with the actual mining and ore development which included stages of concentration to assure optimum metal retrieval. In early America, metallurgists were associated with the mining, milling, and processing of ores to extract raw metals on the one hand, to the blacksmiths throughout the country who, in addition to shoeing horses, produced wagon wheel rims, a great variety of tools (including saws, axes, wedges, hammers, etc.), commodity items such as nails, pots, metal doors, etc. Specialty metal processing for large iron and steel items including gears, machinery and other machinery parts, cannons, locomotives and locomotive parts, rails, and a multitude of industrial metal components and products grew worldwide only over the past several hundred years, driving both the agricultural and industrial revolutions and establishing the basis for all contemporary manufacturing and manufacturing processes. Metallurgists became the professionals who innovated, implemented, and operated nearly every metal process and production component known to modern societies. Metals, more than any other material, created the tools and machinery which have fashioned the evolution of modern civilizations, including the so-called engines of war and the rail networks that founded the transportation and communications networks which today provide global unity.

References

Agricola G (1950) De Re Metallica (First Latin edition 1556) (trans: Hoover HC, Hoover LH). Dover Publications, New York

Examples of Materials Science and Engineering in Antiquity

Contents

Abstract

In this chapter we present a few examples of materials science and engineering and nanomaterials and nanotechnology represented in antiquity. These examples may seem to represent contemporary materials science and engineering concepts but also point up that they are historically pervasive. While ancient metallurgists and materials-related artisans and technologists may not have purposefully devised or designed a materials system, millennia of anecdotal evidence or observations provide compelling support. The examples begin with prospects for concrete casting of some core blocks in the pyramids rather than transporting them from quarries. Asbestos in antiquity illustrates possibly the most multifunctional material ever developed. Carbon nanotubes and related carbonaceous materials are also observed in antiquity along with applications of carbon block and soot derivatives in inks. Nanofibers also play a role in Mayan blue paintings, and colloidal silver and copper are shown to exhibit antibacterial and antifungal action over several millennia linking to contemporary applications involving silver nanoparticles in fiber weaves for socks as a simple example.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_2

Introduction

While the so-called discipline of materials science and engineering did not exist prior to roughly 1960, blends of chemistry, physics (particularly materials chemistry and the physics of materials) along with metallurgy (especially physical metallurgy and process metallurgy) can be found even in antiquity. Notable among these are the extraction and processing of metals and alloys noted above, the discovery and evolution of glass technology, and the substitution of elements as elemental nanoclusters into the irregular (amorphous) silicate structure to produce colors, carbon black nanoaggregate use in inks, colloidal (~30 nm) silver in water as an antibacterial, and nanofiber silicates (clays and asbestos) in a wide variety of structural composites (including stucco and fresco compositions) to name but a few. Many of these examples constitute what are popularly characterized as composite materials, including natural clay which is a mixture of clay minerals and other accessory minerals derived from rocks or sediments. These clay compositions were also often mixed with straw as a reinforcement in the manufacture of building blocks. Similar examples of concrete existed in the Roman empire where volcanic ash and lime mixtures were used in construction of buildings such as the Pantheon.

Concrete Casting and Pyramid Construction

More recent evidence has even suggested that ancient concrete technology was employed at least in part in the construction of the Great Pyramids of Giza (one of the Seven Wonders of the World) by casting the core blocks in situ using granular limestone aggregates and an alkali aluminosilicate-based binder (Barsoum et al. 2006). This process is in contrast to the long prevailing model where limestone blocks averaging 2.5 t were cut in local quarries to precise shapes using copper tools, transported to the pyramid sites, hauled up ramps using thousands of workers, and hoisted in place. The Khufu pyramid (the largest pyramid) itself contains roughly 2.3 million such blocks, many as close as 0.05 mm fit across their contiguous vertical faces, especially the white or gray limestone (marble) casing stones which surfaced the pyramid, having slant faces and flat tops. These were almost entirely removed over the ages, leaving the stepped core structure seen today in the photographs of Fig. 1. The casing stones created a smooth surface resembling other marble structures such as the Taj Mahal. The stepped corner shown in Fig. 1b shows loosened core blocks generally resulting from earthquakes over the millennia since the construction of the pyramids as well as events surrounding the removal of the casing blocks.

The development and use of concrete of various compositions has progressed over many millennia and continues today, mostly as Portland cement compositions (Hewlett 1998). These and precursor cement compositions have been mixed with strong fibers such as chrysotile asbestos nanotube fibers especially in the twentieth century in the fabrication of structural cement members and asbestos-cement pipe in water systems, which in the USA alone could circle the Earth eight times and run

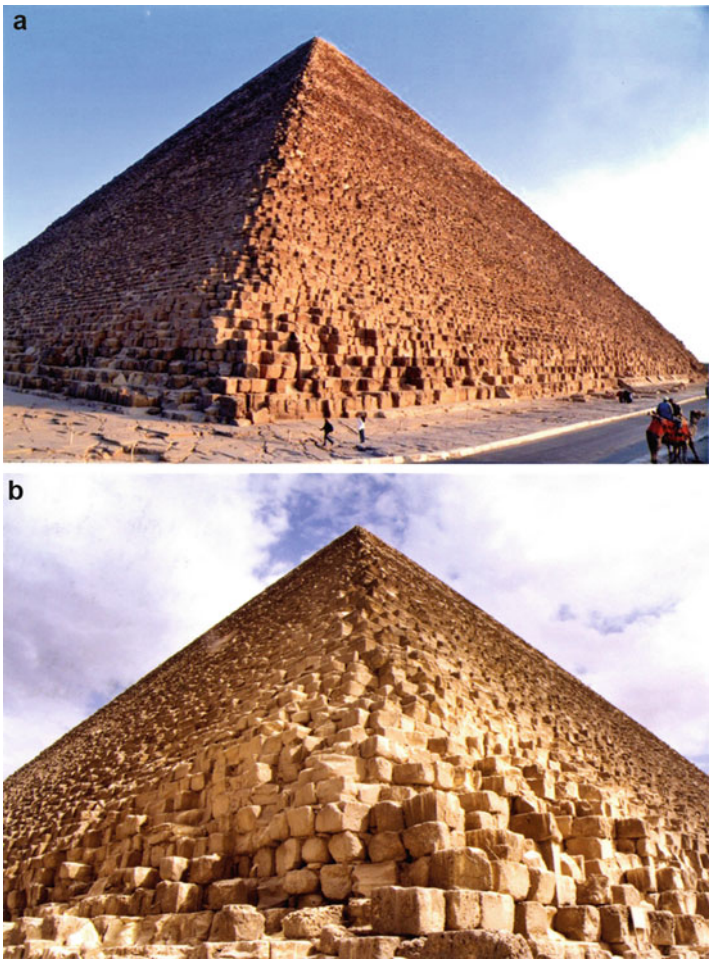


Fig. 1 Khufu pyramid. (**a**) Perspective view. (**b**) Enlarged view showing core blocks along edge base

to the moon and back. These construction materials also included exterior stucco compositions which contained asbestos fibers until health controls were imposed in the later 1900s when lung cancers were shown to be caused by asbestos fiber inhalation. Modern stucco compositions replace asbestos with polymer and other benign and even plant waste fibers for reinforcement.

The Asbestos Millennia: Multifunctional Nanotube Applications

There are roughly 52 different asbestos species-fibrous, hydrated silicates. These hydrated silicate fibers are characterized by amphiboles: nanofiber amosite

($Fe_7Si_8O_{22}$ $(OH)_2$), crocidolite (Na_2 (Fe, Mg)$_5Si_8O_{22}$ $(OH)_2$), tremolite ($Ca_2Mg_5Si_8O_{22}$ $(OH)_2$), actinolite (Ca_2 (Mg, Fe)$_5Si_8O_{22}$ $(OH)_2$) and anthophyllite ((Mg, Fe)$_7Si_8O_{22}$ $(OH)_2$), and serpentine (serpentlike) chrysotile nanotube bundles ($Mg_3Si_2O_5$ $(OH)_4$). Chrysotile has historically accounted for roughly 95 % of asbestos nanofibers used commercially.

Asbestos (especially chrysotile asbestos) has been known since the time of the pyramids as an unusual and often thought magical material since elaborate cloth vestments were fire resistant. Chrysotile asbestos as illustrated in Fig. 2a could be spun into yarn and woven like cotton or wool into expensive textiles. Around 300 BC, the Greeks referred to fire-resistant fabrics as "asbestos," meaning inextinguishable. Health effects were known even in ancient times and the Roman naturalist, Pliny the Elder (circa 70 AD) referred to "the sickness in the lungs" of slaves who worked the asbestos mines. However, it was not until fibrotic lung damage (asbestosis) was documented in English textile workers around 1930 that health regulations were enacted. The USA followed in the 1950s over concerns for shipyard and related workers in the aftermath of World War II. Up to this time, more than 3,000 different uses for chrysotile asbestos can be identified (Alleman and Mossman 1997). Asbestos use declined during the Middle Ages but resumed in parts of Europe in the sixteenth century. The Industrial Revolution of the late 1800s represented a boom for asbestos products which, as noted above, numbered in the thousands by the twentieth century, utilizing millions of tons of asbestos (mainly chrysotile) worldwide annually (Alleman and Mossman 1997).

As shown in Fig. 2b, c, chrysotile asbestos nanotube fibers are composed of curved (cylindrical) crystalline layers separated 7.3 Å. Fibers, which can measure many centimeters in length (Fig. 2a), are often closed (arrow in Fig. 2b) and are nearly indistinguishable from multiwall carbon nanotubes illustrated in Fig. 3.

Carbon Nanotubes: From Antiquity to the Future

The contrast between chrysotile asbestos nanotubes in Fig. 2b and multiwall carbon nanotubes in Fig. 3 is striking. Although carbon nanotubes consist of cylindrical sheets of graphitic (hexagonal) carbon referred to in single sheets as graphene, multiwall carbon nanotubes (MWCNTs) are not crystalline in the sense of chrysotile asbestos. These features are illustrated in Fig. 4 which illustrates the nucleation and growth of MWCNTs from energetically favorable carbon hemispheres containing pentagonal carbon structures which create curvature in the hexagonal (carbon) graphene. Each hemispherical nucleus must contain six pentagons, each ideally contributing ~ 30° curvature. This is in contrast to a closed form fullerene structure which must contain 12 pentagons – such as the C60 buckyball – a truncated icosahedron containing 20 hexagons and 12 pentagons consisting of 60 carbon atoms at each vertex.

Figure 4a shows two so-called armchair hemispherical nuclei growing one over the other and extending as cylindrical carbon tubes, ideally energetically more

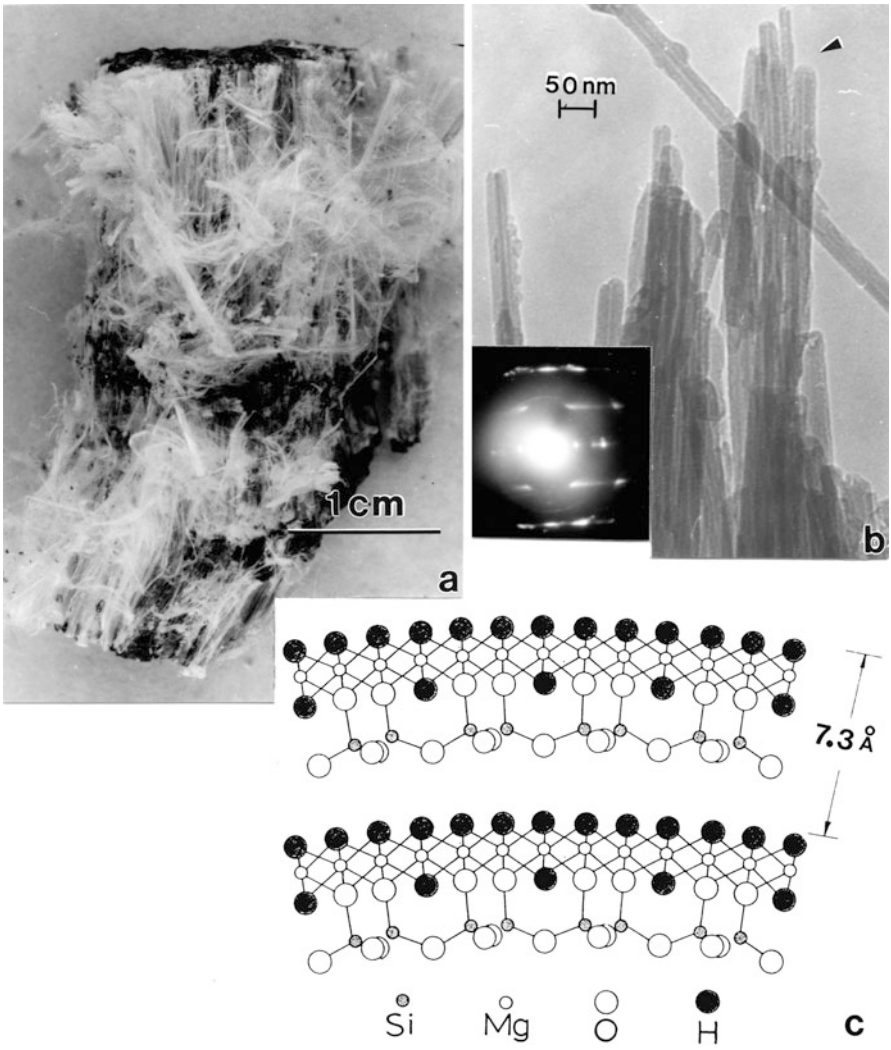


Fig. 2 Chrysotile asbestos. (**a**) Chunk of mined mineral. (**b**) Transmission electron microscope image of individual nanotubes, many with closed ends (*arrow*). The selected-area electron diffraction pattern insert shows diffraction streaks perpendicular to the fiber axes. (**c**) Curved silicate layer structure schematic

favorable than closing off to create a fullerene. As these tubes grow, nonequilibrium events cause them to close, forming varied tube lengths, tube types, and even fullerenic structures as shown in Fig. 3. As shown in Fig. 4a, b, the carbon tube spacing is essentially the same as the equilibrium graphite hexagonal crystal half-cell dimension (~3.4 Å). Figure 4c illustrates this crystal structure in contrast to the cubic diamond structure where the unit cell dimension is 3.6 Å.

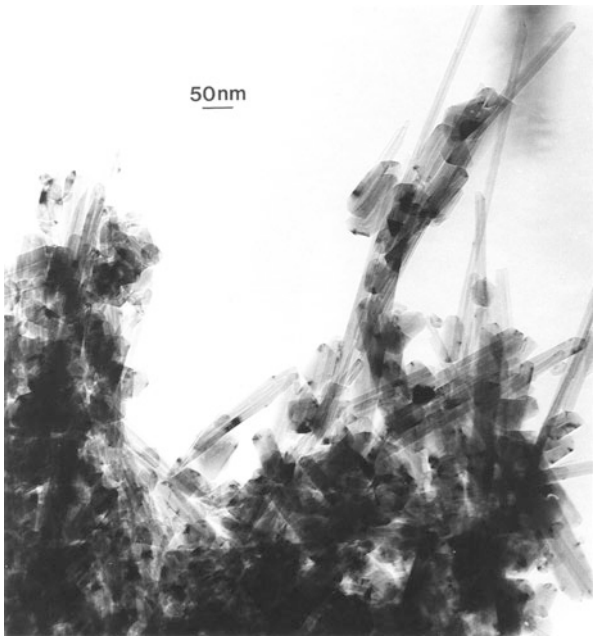


Fig. 3 Transmission electron microscope view showing carbon arc-evaporation-grown multiwall carbon nanotubes and fullerenic nanostructures

Figure 5 shows the conventions for carbon nanotubes in relation to the two-dimensional graphene lattice. This convention identifies the so-called zigzag (Z or type 1) nanotubes along the $\theta = 0°$ reference line and armchair (A or type 2) nanotubes with reference to $\theta = 30°$. Tubes referenced along the vector **C** for $0 < \theta < 30°$ are referred to as chiral. The 2D lattice points (or carbon atom sites in the hexagonal graphene sheet) are noted by n, m, where for Z tubes, $n \neq 0$, $m = 0$. For the A tubes, $n = m$. The tube axis in Fig. 5 is perpendicular to the reference direction, **C**. For tubes with position coordinates, n, m, rolled up and placed on 0, 0, the tube circumference is the scalar value (or length) $|\mathbf{C}|$. Correspondingly, the tube diameter is given by $d = |\mathbf{C}|/\pi = a\,(n^2 + nm + m^2)^{1/2}/\pi$, where $a = 2.46$ Å. Consequently, for the inner (A) tube in Fig. 4(a) where $n = m = 5$, the tube diameter is 6.8 Å (0.68 nm).

In contrast to chrysotile asbestos nanotubes, multiwall carbon nanotubes are also very strong but exhibit either metallic conductivity for Z tubes or semiconducting behavior for A tubes, while chrysotile nanotubes are insulating (nonconducting). While there are no known applications of MWCNTs in antiquity, they have existed for unknown times in antiquity since they have been detected in ice cores dated to 8000 B.C. and presumably trapped from the atmosphere of that period (Murr and Esquivel 2004). Indeed, MWCNTs have been observed as a consequence of combustion processes in modern times, including natural gas (methane, CH_4) combustion, wood burning, etc. (Murr et al. 2004a; Murr and Guerrero 2006). Figure 6a–e shows a common aggregate of nano-silica (SiO_2) particles and MWCNTs collected in the US southwest desert air at the early part of the twenty-first century,

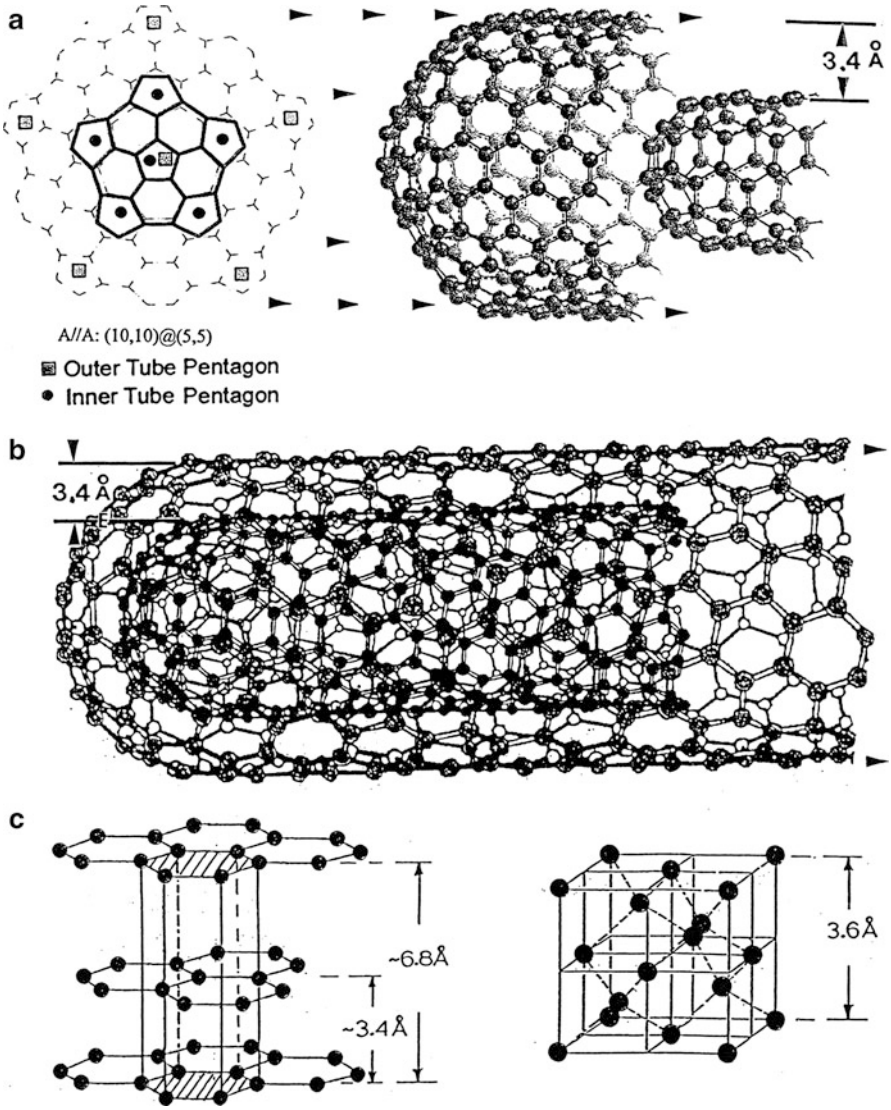


Fig. 4 Carbon nanotube nucleation and growth fundamentals. (**a**) Nucleation and growth of a double-wall carbon nanotube from two overlapping armchair hemispherical caps (A/A) containing curvature-forming pentagons. Computer simulations from Lair et al. (2006). (**b**) Extended double-wall tube growth with tubes separated 3.4 Å. (**c**) Comparison of crystalline carbon unit cells: hexagonal close-packed structure at *left*; diamond cubic unit cell at *right*

while Fig. 6f, g show a similar aggregate observed in a 10,000-year-old Greenland ice core melt (Murr and Esquivel 2004; Murr et al. 2004b). Figure 7 shows a MWCNT-fullerene aggregate collected on a roof top near a natural gas (~96 % methane) water heater exhaust (Murr 2008).

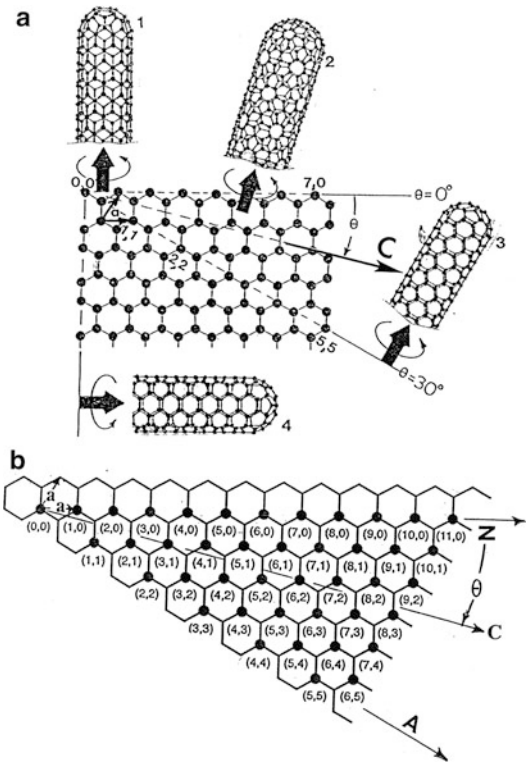


Fig. 5 (**a**) Carbon (graphene) lattice construction and conventions for characterizing carbon nanotube types. (*1*) and (*2*) show zigzag and chiral tubes, respectively, with tube axis perpendicular to *C*. (*3*) and (*4*) are armchair tubes. (**b**) shows the indexed graphene lattice where the lattice vectors **a** have scalar values $|\mathbf{a}| = 2.46$ Å. Zigzag (*Z*) tubes are perpendicular to *C*: $\theta = 0$, $n \neq 0$, $m = 0$. Armchair (*A*) tube axis is perpendicular to *C* for $\theta = 30°$, $n = m$

Beginning with the start of the twenty-first century, applications of carbon nanotubes and single or controlled layers of graphene were under intensive study, and additions of carbon nanotubes to a variety of matrix materials to provide composite strength were common. Although some in vitro and in vivo studies of MWCNTs in particular have raised some health issues, the prospects for longer-term chronic health effects (including respiratory cancers) like asbestos were not sufficiently documented by 2012.

Black Carbon and Related Combustion Soots

Like MWCNTs and fullerenes (which have been detected in geological deposits dating at least a million years ago (Buseck et al. 1992), and uncontrolled combustion soots date well before the emergence of humans, while various forms of black carbons (BC) found uses in paints and inks dating several millennia B.C. (Barabe et al. 2006; Carvalho 1904). Unlike carbon nanotubes discussed briefly above, BC nanoparticulates were used in ancient times. BC, unlike more common combustion soots, is primarily composed of irregular (turbostratic) curved graphene

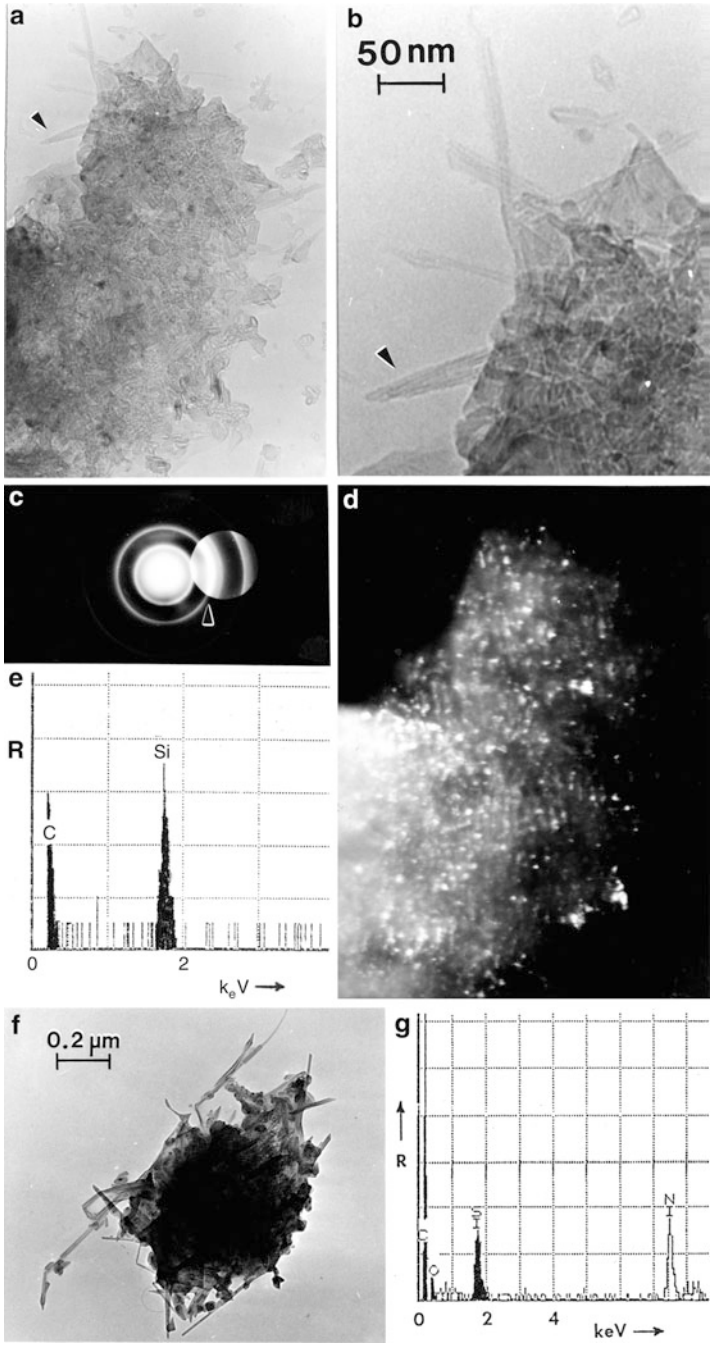


Fig. 6 Comparison of nano-silica/multiwall carbon nanotube aggregates by transmission electron microscopy. (**a**) TEM view of atmospheric carbon nanotube aggregate. (**b**) Magnified view of (**a**)

Fig. 7 Multiwall carbon nanotube aggregate collected near a natural gas hot water heater home roof exhaust (by thermal precipitation) and observed in the transmission electron microscope

(C) fragments composing spherical particulates 20–40 nm diameter which are usually aggregated. Soots, on the other hand, appear similar, but sometimes as large as 80 nm single particles. In contrast, soot particulates are composed of the same curved graphene fragments intercalated with polycyclic aromatic hydrocarbon (PAH) molecules – C_xH_y – based on benzene (C_6H_6), with a molecular weight (MW) of 78. These become higher molecular weight compounds (isomers) depending upon the combustion thermo-kinetics. Figure 8 shows several views for a soot aggregate for a burning tire, where the principal, intercalated PAH molecules include (Shi et al. 2000) anthracene ($C_{14}H_{10}$, MW = 178), phenanthrene

Fig. 6 (continued) at *arrow* showing multiwall carbon nanotube. (**c**) Selected-area electron diffraction pattern of (**a**). (**d**) Dark-field TEM image of (**a**) using the selected aperture position in (**c**) at *arrow* showing silica (SiO_2) nanoparticles in the aggregate. (**e**) Energy-dispersive x-ray spectra for (**a**) showing carbon and silicon peaks. (**f**) TEM image showing a similar multiwall carbon nanotube-silica aggregate in a 10,000-year-old Greenland ice core. (**g**) Energy-dispersive x-ray spectrum similar to (**e**) and corresponding to (**f**). Note C, O, and Si peaks. Ni peak shown is reference coating used for calibration (Adapted from Murr et al. 2004b)

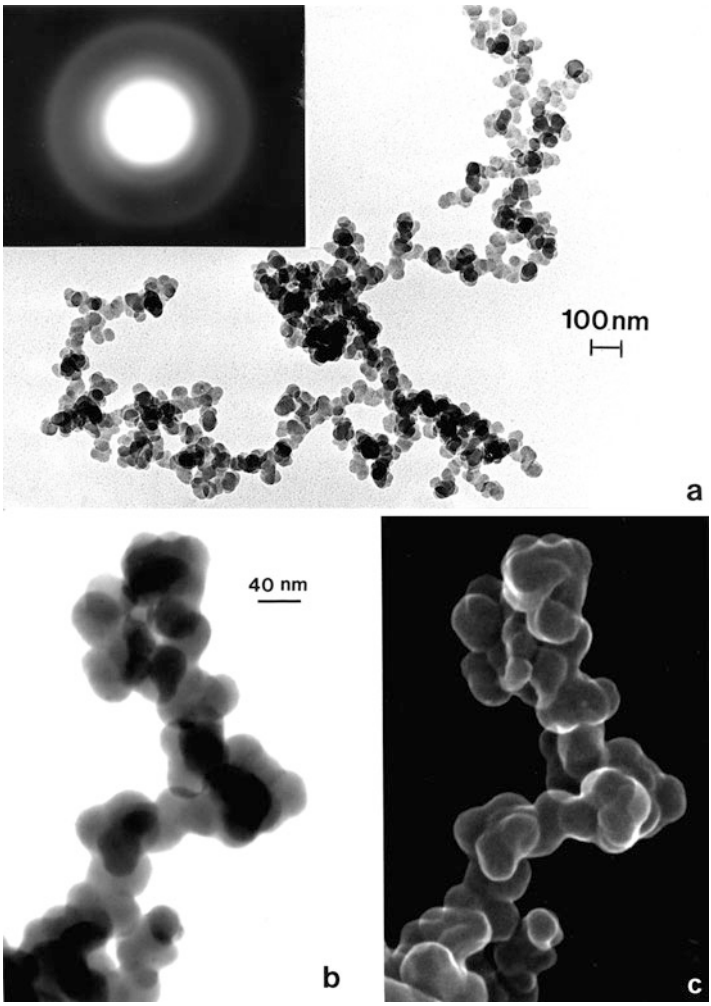


Fig. 8 Example of branched, aggregated soot nanospheres collected from a burning tire and observed in electron microscopes. (**a**) Transmission electron microscope image with diffuse graphite selected-area electron diffraction pattern insert. (**b**) Scanning electron microscope view in the scanning electron transmission mode. (**c**) Secondary electron image of (**b**)

($C_{14}H_{10}$, MW = 178), fluoranthene ($C_{16}H_{10}$, MW = 202), pyrene ($C_{16}H_{10}$, MW = 202), chrysene ($C_{18}H_{12}$, MW = 228), and molecular weights up to 278. PAHs with the same molecular weight have different molecular structures.

Of course in more modern times, BC has been an integral component of black rubber products such as automobile tires which contain roughly 30 % BC as a strengthening agent for the organic (polymeric) rubber matrix. Currently more than five million tons of BC are used commercially in the manufacture of products in the USA alone. BC also continues to be employed in black ink manufacture.

Nanofibers in "Ageless" Mayan Paintings

The Maya, the best-known Mesoamerican civilization which rose to prominence around 250 AD in the Yucatan region of Central America (including southern Mexico and Guatemala), built pyramid-like structures in humid jungles adorned by murals whose vivid paint colors have hardly faded over centuries. Particularly notable has been the unusual stability of Mayan blue paint over more than eighteenth centuries in some locations. In contrast to modern-day paints composed of pigments mixed with TiO_2 (titania) white powder (often nanoparticulate aggregates) in a water-based or volatile organic binder which degrades in only a few years, the "ageless" Mayan blue paint has been shown to be composed by indigo dye molecules trapped in zeolite-like channels in silicate nanofibers measuring 3.4 × 6.4 Å. This inorganic, palygorskite ($(Mg, Al)_4\ Si_8\ (O, OH, H_2O)_{24}$) silicate mineral fiber host (an asbestiform mineral) forms silicate clay sheets with organic indigo molecules ($C_{16}H_{10}H_2O_2$) locked in channels near the clay surfaces by hydrogen bonds between indigo carbonyl and structural molecules (Fois et al. 2003; Polette et al. 2000).

Figure 9 provides a visual summary of the remarkable Maya blue paint. Figure 9a shows an enlarged section from a large wall mural painted on a dry stucco base in ancient Cacaxtla (Bonampak), near Mexico City, around 850, near the end of the Mayan era. Figure 9b shows a transmission electron microscope image of palygorskite fiber bundles, while Fig. 9c illustrates a computed, high-resolution transmission electron microscope image for a fiber end showing the sorbed indigo molecule illustrated in Fig. 9d. In contrast to the palygorskite nanofibers central to Mayan blue paint, Fig. 9e shows TiO_2 nanoparticles and particle aggregates observed in the transmission electron microscope which are used in contemporary white base paint. The palygorskite clay sheets, when pulled apart as shown in Fig. 9b, not only locked the blue indigo dye molecule but in the applied paint binder, which often included sticky tree saps, provided the strength afforded by strong fibers like chrysotile asbestos. Indeed, Maya blue paint in particular epitomizes modern hybrid materials characterized by compounds of inorganic and organic design having unusual properties and functionality.

Medical and Antimicrobial Uses of Copper and Silver

Although not specifically in the province of materials science and engineering, it is of interest to observe the creation of medical devices made from copper and silver and their use based largely on the anecdotal evidence that they reduced or prevented infection. The oldest medical uses of copper were recorded in an Egyptian medical text written around 2400 BC called the Smith Papyrus, considered one of the oldest known books (Dollwet and Sorensen 1985). These uses included the fabrication of vessels to sterilize drinking water and the pressing of copper objects against wounds to sterilize them. Similar applications were made for silver, which, like copper, was first exploited in native forms and later extracted by heating geological ores. Both copper and silver urns were cast or hammered to form by the Egyptians, Greeks,

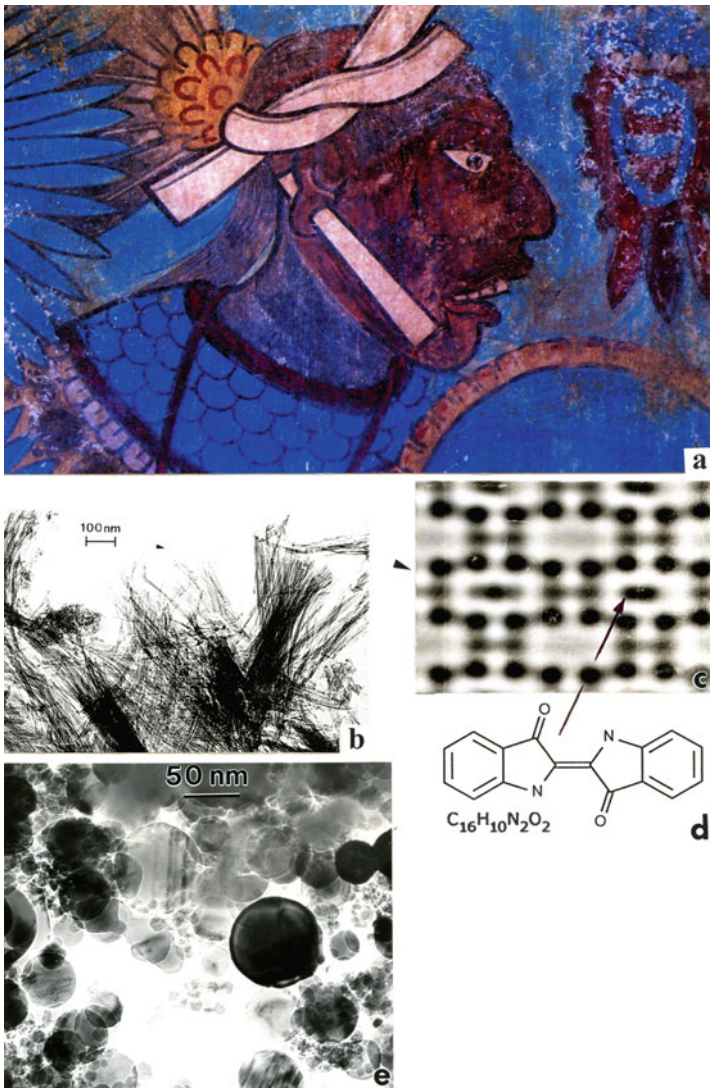


Fig. 9 Visualizing ancient Mayan blue paint. (**a**) Enlarged section of later-Mayan mural in Cacaxtla (Mexico). (**b**) Transmission electron microscope image of palygorskite nanofiber bundles. (**c**) Computed image for palygorskite fiber end showing trapped indigo molecule (**d**) in channels. (**e**) Transmission electron microscope image for titania (TiO_2) nanoparticulates forming white paint base. (**d**) is courtesy of R. R. Chianelli

Romans, and many other civilizations for water and other liquid storage and purification. Correspondingly both copper and silver appliances and compounds of copper and silver were used for burns and various wounds. Hippocrates, the Greek physician often considered the father of modern medicine, wrote that silver

could be used to prevent disease and had special healing power (D'Raye 1998). In the Middle Ages silverware of the wealthy was recognized to provide some protection from the plague, and wealthy children were given silver spoons to suck on to prevent plague. Plague or "Black Death" has been documented from DNA analysis to have been caused by the *Yersinia pestis* bacterism probably carried by rats or rat fleas. This use of silver spoons by wealthy children gave rise to the expression, "born with a silver spoon in his mouth." In this context, ancient Chinese Emperors were known to eat with silver chopsticks as a hygienic precaution and for good health. Silver surgical instruments were popular in many parts of the world for similar reasons from at least the sixteenth century.

Figure 10a shows one of the two largest sterling silver vessels in the world in Jaipur, Rajasthan, India, fabricated at the turn of the twentieth century for Hindu King Maharaja Sawai Madho Singh II. Their purpose was to allow the king to take sacred Ganges River water on his travel to England in 1901. The 92.5 % Ag, 7.5 % Cu sterling silver vessel in Fig. 10 is 1.7 m high and weighs 340 kg when empty. With a 4,000 l capacity, it would add 4,000 kg when filled with Ganges water. Each vessel was cast from 14,000 melted silver coins and added copper to produce the sterling silver alloy. Silver coins themselves were also often thrown into wells or water barrels on ships in order to kill bacteria or simply to keep the water "fresh."

Silver has dominated as the metal of choice for antimicrobial and antiviral medicinals (D'Raye 1998), and for several hundred years, colloidal silver (with particle diameter of ~ 30 nm as shown in Fig. 10b) in concentrations of 10–20 ppm in purified water has found wide acceptance and is available in pharmacies in many countries. Silver nanoparticles are currently woven into hospital fabrics and many other fabrics such as socks to eliminate bacteria and fungi.

The use of both copper and silver as well as copper alloys (brass in particular: 70 % Cu-30 % Zn) and sterling silver as self-sanitizing surfaces, especially in hospital environments and other hygiene-sensitive areas, is becoming increasingly popular. The emergence of antibiotic-resistant bacteria, especially ubiquitous in hospitals, nursing homes, and food processing facilities around the world, has led to wide-ranging uses of copper, copper alloys, and sterling silver surfaces for contact-mediated killing of bacteria (Crannel 1967; Grass et al. 2011). Neither the killing of pathogens or viruses in vivo nor the contact killing of bacteria on metal or alloy surfaces involving copper or silver is well documented mechanistically. Colloidal Cu or Ag as illustrated in Fig. 10b does not appear to bind directly to microbes, and sterilization in vessels presents another regime. More generally, reactive oxygen species (ROS) may be mediated by redox cycling between the neutral metal (M(0)) and the two valence states +1 and +2: M(+1) and M(+2). For Cu the valence electrons are characterized by $3d^{10}4s^1$, while for Ag the valence electrons are characterized by $4d^{10}5s^1$. These intervalence transfers in ionic forms of the metals can create peroxide (H_2O_2) and extremely reactive hydroxyl radical:

$$2M^+ + 2H^+ + O_2 \rightarrow 2M^{2+} + H_2O_2$$

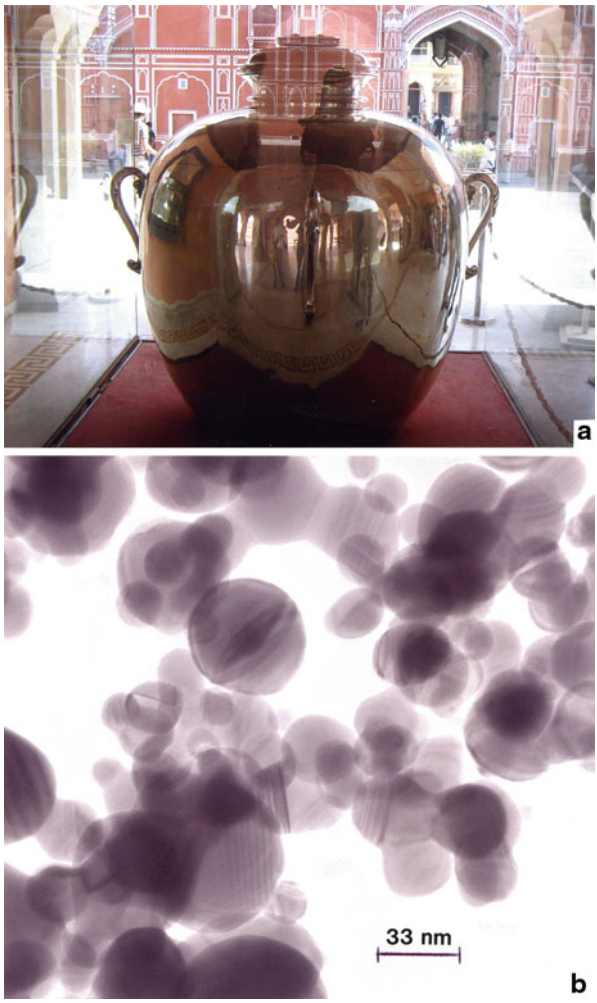


Fig. 10 (**a**) Person-size sterling silver vessel in Jaipur, India, fabricated for Ganges river water storage and sterilization around 1900. (**b**) Transmission electron microscope image of silver nanoparticulate aggregate

$$M^{+} + H_2O_2 \rightarrow M^{2+} + OH^{-} + OH$$

where M ⇒ Cu or Ag. It is also possible that M ions cause cell damage directly by cell membrane rupture (Grass et al. 2011; Warnes et al. 2010; Parsons et al. 2005).

References

Alleman JE, Mossman BT (1997) Asbestos revisited. Sci Am 7:70–75

Barabe JG, Martin KA, Schumacher EF, Swider JR, Teetsov AS (2006) Examination of the Gospel of Judas using an integrated approach to ink characterization. Microsc Today 14(4):6–15

Barsoum MW, Ganguly A, Hug G (2006) Microstructural evidence of reconstituted limestone blocks in the Great Pyramids of Egypt. J Amter Ceram Soc 89(12):3788–3796
Buseck RR, Tsipursky SJ, Hettich R (1992) Fullerenes from the geological environment. Science 257(5067):215–217
Carvalho DN (1904) Forty centuries of ink. B. Franklin, New York, pp 63–81
Crannel MY (1967) Chapter 17: Silver in medicine. In: Butts A (ed) Silver, economics, metallurgy and use. Van Nostrand, Princeton, pp 227–237
D'Raye T (1998) The wonder cure time forgot: colloidal silver, Revisedth edn. AWIECA, London
Dollwet HHA, Sorenson JRJ (1985) Historic uses of copper compounds in medicine. Trace Elem Med 2:80–87
Fois E, Gamba A, Tilocca A (2003) On the unusual stability of Maya blue paint: molecular dynamics simulations. Microporous Mesoporous Mater 57:263–272
Grass G, Rensing C, Solioz M (2011) Metallic copper as an antimicrobial surface. Appl Environ Microbiol 77(5):1541–1547
Hewlett PC (ed) (1998) Lea's chemistry of cement and concrete, 4th edn. Arnold Publishing, New York
Lair SL, Herndon WC, Murr LE (2006) End cap nucleation of carbon nanotubes. Carbon 44:447–455
Murr LE (2008) Microstructures and nanostructures for environmental carbon nanotubes and nanoparticulate soots. Int J Environ Res Public Health 5(5):321–336
Murr LE, Esquivel EV (2004) A TEM analysis of nanoparticulates in a polar ice core. Mater Charact 52(1):15–25
Murr LE, Guerrero PA (2006) Carbon nanotubes in wood soot. Atmos Sci Lett 7:93–95
Murr LE, Bang JJ, Esquivel EV, Guerrero PA, Lopez DA (2004a) Carbon nanotubes, nanocrystal forms, and complex nanoparticle aggregates in common fuel-gas combustion sources and the ambient air. J Nanoparticle Res 6:241–251
Murr LE, Esquivel EV, Bang JJ, de la Rosa G, Gardea-Torresdey H (2004b) Chemistry and nanoparticle compositions of a 10,000 year-old ice core melt water. Water Res 38(19):4282–4296
Parsons D, Bowler PG, Myles V, Jones S (2005) Silver antibacterial, physical and chemical characteristics. Wounds 17(8):222–232
Polette LA, Ugarte N, Yacaman MJ, Chianelli RR (2000) Decoding the complexity of a remarkable ancient paint. Discov Archeol Sci Amer (July–August) 7:46–53
Shi Y, Murr LE, Soto KF, Lee WY, Guerrero PA, Ramirez DA (2000) Characterization and comparison of speciated atmospheric carbonaceous particulates and their polycyclic aromatic hydrocarbon contents in the context of the Paso Del Norte airshed along the U.S.-Mexico border. Polycycl Aromat Comp 27:361–400
Warnes SL, Green SM, Michels HT, Keevil CW (2010) Biocidal efficacy of copper alloys against pathogenic enterococci involves degradation of genomic and plasmid DNA. Appl Environ Microbiol 76:5390–5401

Part II

Electricity and Electromagnetic Phenomena: The Historical Perspective for Materials Fundamentals

Electromagnetic Fundamentals

Contents

Abstract

Following a brief overview or review of the history of electricity, electromagnetic waves and related phenomena are discussed in the context of Maxwell's equations and the development of the so-called wave-optics equation which was the precursor to the Schrödinger equation as will be discussed in the chapter "► A Brief Introduction to Quantum Mechanics." The important issues to be illustrated in this chapter (chapter "► Electromagnetic Fundamentals") are the identification and evolution of fundamental materials properties: permittivity (ε), permeability (μ), index of refraction (n), conductivity and resistivity (σ and ρ), Ohm's law, and the interrelationship of these materials properties as well as their measurement. This provides a context for materials properties historically and illustrates the role that Maxwell and his mathematical contributions have had in establishing quantum mechanics and its fundamental role in the behavior of materials, particularly nanomaterials.

Introduction

It is interesting to realize that the electric and magnetic properties of materials were essentially the first materials properties to be somewhat systematically exploited in the context of materials science and engineering. Most often, materials science and engineering fundamentals initially explore atomic or electronic structure as this

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_3

relates to crystal structure. But long before chemical elements were categorized in the context of the periodic table, fundamental properties of materials were beginning to be identified as early as the seventeenth century. Although, as noted in chapters "▶ A Brief History of Metals" and "▶ Examples of Materials Science and Engineering in Antiquity," materials, even nanomaterials, were utilized for their performance and were processed in antiquity, specific properties involving these performance issues were not specifically codified, with the exception of hardness and ductility which were almost intuitively recognized to differentiate materials, especially metals and alloys as these were discovered and developed over several millennia.

This chapter provides a rather nontraditional basis for materials and especially electric and magnetic materials and materials properties. In this context, fundamental properties of materials will be presented: permittivity (ε) characterizing dielectric materials, permeability (μ) characterizing magnetic materials, and index of refraction as it relates to these properties and describes the velocity of light in a material, especially transparent materials. Conductivity and current density will be discussed in the context of the four-point forms of Maxwell's equations as well as the development of the electromagnetic wave equation which relates the index of refraction (n) and the velocity of light (c). This chapter provides an overview of electromagnetic waves quantized as component photons originally conceptualized by A. Einstein to account for the interaction (absorption and emission) of the visible electromagnetic spectrum with matter. These phenomena lead to color in materials to be discussed in the chapter "▶ Electromagnetic Color and Color in Materials."

The history of electricity began with the publication of a Latin treatise called De Magnete, Magneticisque Corporibus (On the Magnet) by W. Gilbert around 1600 explaining his experiments with both electricity and magnetism, wherein he coined the word "electrica." In 1729, Stephen Gray discovered electric conductivity, and 4 years later, C. F. duFay discovered that electricity may be positive or negative in polarity. A few years later, in 1747, Henry Cavendish was the first to measure conductivity (σ) for numerous materials (particularly metals such as copper, gold, and silver). About this time, Pieter van Musschenbroek invented the Leyden jar, the first capacitor for storing electricity, and made famous by Ben Franklin's kite experiment a few years later (Agassi 2007).

Around 1825–1830, A. Volta and A. Ampere, along with G. S. Ohm, demonstrated the connection between the "volt," "ampere," and "ohm," where the practical unit of electric current strength (I) was given with an electromotive force of 1 V (V) through a conducting wire having a resistance (R) of 1 Ω. This gave rise to Ohm's law: volts = amperes × resistance or $V = IR$ ($1\ V = 1\ A \times 1\ \Omega$). The ohm is often expressed by Ω.

In 1831, J. Henry discovered the essential mechanics for the electric telegraph invented by Samuel Morse a year later. Joseph Henry and M. Faraday codiscovered the dynamo or electric generator principle around this time period wherein a conducting loop moving in an appropriately oriented magnetic field created an electromotive force and current flow in the conducting wire loop, characterized

fundamentally as Faraday's law. These inventions led to arc lighting for streetlights in 1878 by C. Brush just a few years before Thomas Edison developed the incandescent lamp using a carbonized filament powered by direct current (dc). This was followed by the development of alternating current (ac) systems by George Westinghouse who bought the patent rights to these concepts of ac systems developed by Nikola Tesla in 1885. Westinghouse used these ac systems to light the Columbian Exposition in Chicago in 1893.

When an electric field exists in a conducting material such as copper, which continues to be the universal electric carrier, electric charge (Q) will flow in the direction of the electric field (E). This charge flow, or electric current (I) through any given surface or surface area (da), is equal to the rate of flow of electric charge past that surface, or

$$I = dQ/dt \tag{1}$$

where Q is the total charge passing through the surface. Additionally, by Ohm's law

$$I = V/R, \tag{2}$$

where $1/R = \sigma$ (area/length) and the conductivity $\sigma = 1/\rho$, where the ρ is the resistivity (in units of ohm-cm). Voltage can be defined by

$$V = \int \mathbf{E} \cdot ds \tag{3}$$

and correspondingly

$$I = \int \mathbf{J} \cdot da \tag{4}$$

where **E** and **J** are the electric field intensity and current density, respectively, and d**s** and d**a** are described in the wire section represented in Fig. 1 and considered to be vectors. The current flowing through this section is $\mathbf{J} \cdot d\mathbf{a}$, while the voltage along the section (d**s**) is $\mathbf{E} \cdot d\mathbf{s}$. Then

$$\mathbf{J} \cdot d\mathbf{a} = \sigma(da/ds)\mathbf{E} \cdot d\mathbf{s}, \tag{5}$$

whereupon current density and electric field intensity (or field strength) are related by

$$\mathbf{J} = \sigma\mathbf{E} \tag{6}$$

which is popularly referred to as a microscopic Ohm's law. For an isotropic, homogeneous material, σ is a scalar, but it is a tensor, $\boldsymbol{\sigma}$, if the material behaves anisotropically. **J** and **E** are both vector quantities which are colinear as shown in Fig. 1.

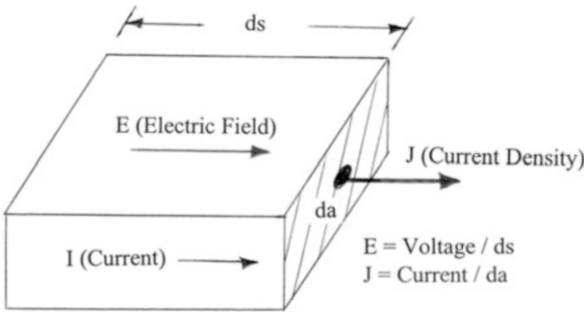


Fig. 1 Electric conductor section

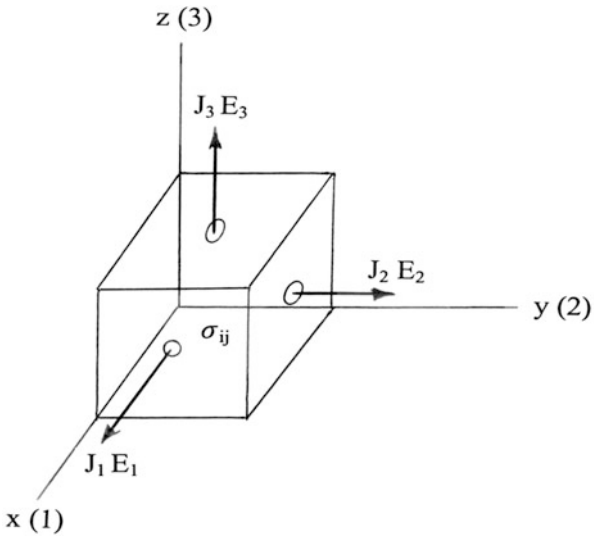


Fig. 2 Conducting element in a Cartesian coordinate system

In tensor notation

$$\mathbf{J} = \boldsymbol{\sigma}\mathbf{E} \tag{7}$$

since **J** and **E** are first-rank tensors (vectors), while **σ** is a second-rank tensor. For a 3D solid or solid section (ideally anisotropic), we can write

$$\begin{pmatrix} J_1 \\ J_2 \\ J_3 \end{pmatrix} = \begin{pmatrix} \sigma_{11}\sigma_{12}\sigma_{13} \\ \sigma_{21}\sigma_{22}\sigma_{23} \\ \sigma_{31}\sigma_{32}\sigma_{33} \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix}, \tag{8}$$

where 1, 2, 3 ⇒ X, Y, Z in a Cartesian coordinate system shown ideally in Fig. 2.

Equation 7 can also be written in index notation:

$$J_i = \sigma_{ij}E_j, \tag{9}$$

where i, j ⇒ 1, 2, 3 and is a shorthand expression noted in Eq. 8 above. These concepts will be expanded upon later when tensor notation is revisited in dealing comprehensively with materials properties in anisotropic materials. Consequently, the σ_{ij} composing Eq. 8 represent scalar coefficients of conductivity.

Electromagnetics and Maxwell's Equations

Equation 6 as it applies to isotropic matter is often considered to represent a primary constitutive relation which describes a medium's properties and effects when two physical quantities are related. In electromagnetics, Eq. 6 is one of four constitutive relations:

$$\begin{cases} \mathbf{J} = \sigma\mathbf{E} \\ \mathbf{D} = \varepsilon\mathbf{E} \\ \mathbf{B} = \mu\mathbf{H} \\ \mathbf{M} = \chi_m\mathbf{H} \end{cases}, \tag{10}$$

where the first two describe the relationship between the electric field, **E**, and the current density, **J**, and the electric flux density or electric displacement, **D**, and the other two relate the magnetic flux density of magnetic induction, **B**, and the magnetic field intensity, **H**, and magnetic polarization, **M**. In Eq. 10, σ (the electric conductivity), ε (the dielectric permittivity), μ (the magnetic permeability), and χ_m (the magnetic susceptibility) represent isotropic materials properties; $\chi_m = (\mu_r - 1)$, where μ_r (the relative permeability) $= \mu/\mu_o$. μ_o represents the permeability in so-called free space and is equal to $4\pi \times 10^{-7}$ H/m (Henries/m). Correspondingly, we can define an electric susceptibility: $\chi_e = (\varepsilon_r - 1)$, where ε_r (the relative permittivity) $= \varepsilon/\varepsilon_o$; ε_o represents the permittivity of free space and is equal to 8.85×10^{-12} F/m (Farads/m).

In electromagnetism, permittivity (ε) measures the resistance encountered when an electric field is formed in a material or how an electric field (E) is affected by a dielectric medium. Correspondingly, permeability (μ) describes how matter interacts with a magnetic field (H). For weak magnetic attraction, μ is described as paramagnetic. For strong interaction, μ is ferromagnetic. When atomic spins are aligned, the magnetic polarization is opposite to the magnetic field (−), and μ is referred to as diamagnetic. In free space (vacuum), $\varepsilon = \varepsilon_o$, $\mu = \mu_o$, and ε_r and $\mu_r = 1$. Consequently, χ_e and χ_m are zero (Moliton 2007).

As noted earlier, A. Ampere and M. Faraday made major contributions to electromagnetic phenomena. These are illustrated schematically in Fig. 3 which represents Ampere's law and Faraday's law. In Ampere's law (Fig. 3a), an electric current, I, in a conductor generates an induced magnetic field around it: the right-hand rule where the fingers of the right hand represent this field with the thumb pointing in the current direction. Mathematically, Ampere's law can be expressed as

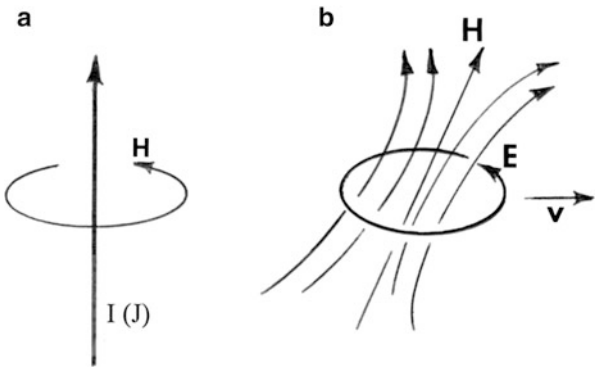


Fig. 3 Illustration of Ampere's law (**a**) and Faraday's law (**b**)

$$\oint \mathbf{H} \cdot \mathrm{d}\mathbf{s} = \mathrm{I} \tag{11}$$

where the line integral represents a closed path. From Eq. 4,

$$\oint \mathbf{H} \cdot \mathrm{d}\mathbf{s} = \int \mathbf{J} \cdot \mathrm{d}\mathbf{a}. \tag{12}$$

Invoking the Stokes' theorem, the left-hand member of Eq. 12 can be changed from a line integral around a closed path to a surface integral over the same surface bounded by that path:

$$\oint \mathbf{H} \cdot \mathrm{d}\mathbf{s} = \int (\nabla \times \mathbf{H}) \cdot \mathrm{d}\mathbf{a} = \int \mathbf{J} \cdot \mathrm{d}\mathbf{a} \tag{13}$$

and

$$\nabla \times \mathbf{H} = \mathbf{J}, \tag{14}$$

where ∇ (defined as "del" or "nabla") is

$$\nabla = \mathbf{i}\frac{\partial}{\partial \mathrm{x}} + \mathbf{j}\frac{\partial}{\partial \mathrm{y}} + \mathbf{k}\frac{\partial}{\partial \mathrm{z}}, \tag{15}$$

a vector operator corresponding to Cartesian coordinates, X, Y, Z. The cross product of Eq. 14 represents the "curl" of the magnetic field intensity, **H**, where the determinant form can be expressed as

$$\nabla \times \mathbf{H} = \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial \mathrm{x}} & \dfrac{\partial}{\partial \mathrm{y}} & \dfrac{\partial}{\partial \mathrm{z}} \\ \mathrm{H_x} & \mathrm{H_y} & \mathrm{H_z} \end{pmatrix}. \tag{16}$$

Correspondingly, Eq. 14 implies that the curl of the magnetic field at any point is proportional to the current density at that point.

Since, in viewing Figs. 1 and 3a, one line of flux emanates from each unit of charge (Q), the rate of increase of flux passing through a surface (da) is

$$\frac{\mathrm{d}\,(\mathrm{flux})}{\mathrm{dt}} = \frac{dQ}{dt} = \mathrm{I}. \tag{17}$$

However,

$$(\mathrm{flux}) = \int \mathbf{D} \cdot \mathrm{d}\mathbf{a} \tag{18}$$

so

$$\mathrm{I} = \frac{d}{dt}\oint\mathbf{D} \cdot \mathrm{Da} = \oint\frac{\partial\mathbf{D}}{\partial\mathbf{t}} \cdot \mathrm{d}\mathbf{a}. \tag{19}$$

But the current flowing out through a closed surface is given by Eq. 4 expressed as a closed integral:

$$\oint\mathbf{J} \cdot \mathrm{d}\mathbf{a},$$

and this will be equal to the current entering. Therefore,

$$\oint \quad \mathbf{J} \cdot \mathrm{d}\mathbf{a} = \oint\frac{d\boldsymbol{D}}{dt} \cdot \mathrm{d}\mathbf{a}, \tag{20}$$

or

$$\oint\left(\mathbf{J} + \frac{\mathrm{d}\mathbf{D}}{\mathrm{dt}}\right) \cdot \mathrm{d}\mathbf{a} = 0. \tag{21}$$

Invoking Gauss's theorem,

$$\oint\mathbf{D} \cdot \mathrm{d}\mathbf{a} = \int \nabla \cdot \mathbf{D}\,\mathrm{dv} \tag{22}$$

Equation 21 becomes

$$\int \nabla \cdot \left(\mathbf{J} + \frac{\mathrm{dD}}{\mathrm{dt}}\right)\mathrm{dv} = 0. \tag{23}$$

Since Eq. 23 is true for space contained within any closed surface large or small, it follows that

$$\nabla \cdot \left(\mathbf{J} + \mathrm{E}\frac{\mathrm{d}\mathbf{D}}{\mathrm{dt}} \right) = 0 \tag{24}$$

or by substitution of Eqs. 6 and 10:

$$\nabla \cdot \left(\sigma \mathbf{E} + \varepsilon \frac{\mathrm{dE}}{\mathrm{dE}} \right) = 0. \tag{25}$$

The implication of Eqs. 24 and 25 is that if the electric field, **E**, is time varying, there will be another current in the medium known as the displacement current. Correspondingly, there will be another contribution to induce the magnetic field, **H**: $\partial \mathbf{D}/\partial \mathrm{t}$. This displacement current works like the conductive current, **J**, so the total current in Eq. 14 should be

$$\nabla \times \mathbf{H} = \mathbf{J} + \partial \mathbf{D}/\partial \mathrm{t}. \tag{26}$$

Similarly, if we examine the mathematical implications for Faraday's law shown schematically in Fig. 3b, the induced electromotive force can be expressed as

$$\oint \mathbf{E} \cdot \mathrm{d}\mathbf{s} = -\frac{\partial Ø}{\partial t}, \tag{27}$$

where Ø is the flux passing through the surface bounded by the conductor, measured in webers, while the electromotive force is measured in volts. By invoking Stokes' theorem,

$$\oint \mathbf{E} \cdot \mathrm{d}\mathbf{s} = \int (\nabla \times \mathbf{E}) \cdot \mathrm{d}\mathbf{a} \tag{28}$$

Equation 27 becomes

$$\int (\nabla \times \mathbf{E}) \cdot \mathrm{d}\mathbf{a} = -\int \frac{\partial \mathbf{B}}{\partial \mathrm{t}} \cdot \mathrm{d}\mathbf{a}, \tag{29}$$

or

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial \mathrm{t}}, \tag{30}$$

which shows that an electric field will have curl in a region where the magnetic field is time varying. In other words, a changing magnetic field (or flux density, **B**) induces an electric field (or field intensity, **E**) in a conductor.

While Eqs. 26 and 30 express the magnetic field in terms of the force exerted on a current-carrying conductor and that a changing magnetic field induces an electric field, neither the source of the magnetic field nor its configuration in space has been considered. In this regard, the summation of the magnetic field (or magnetic flux density, **B**) over every closed surface is zero:

Fig. 4 James Clerk Maxwell (1831–1879)

$$\oint \mathbf{B} \cdot d\mathbf{a} = 0. \tag{31}$$

Applying Gauss's theorem expressed in Eq. 22, Eq. 31 becomes

$$\nabla \cdot \mathbf{B} = 0 \tag{32}$$

which means that the magnetic field has no divergence under any circumstances.

Similarly, electric flux on electrostatic flux is described by

$$\oint \mathbf{D} \cdot d\mathbf{a} = Q \tag{33}$$

where Q is the charge within a closed surface. Again, applying the Gauss theorem, Eq. 33 can be expressed as

$$\int (\nabla \cdot \mathbf{D}) dv = Q = \int \rho dv, \tag{34}$$

where q_e is the charge/unit volume and

$$\nabla \cdot \mathbf{D} = q_e. \tag{35}$$

Collecting Eqs. 29, 30, 32, and 35, these constitute the original mathematical analysis representing the physical principles characterizing electromagnetic problems formulated by James Clerk Maxwell (Fig. 4) and describing electric and magnetic phenomena in materials:

$$\left.\begin{array}{l} \nabla \times \mathbf{H} = \quad \mathbf{J} + \partial \mathbf{D}/\partial t \\ \nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t \\ \nabla \cdot \mathbf{B} = 0 \\ \nabla \cdot \mathbf{D} = q_e \end{array}\right\}. \tag{36}$$

When applied in a homogeneous medium where there is no charge or conductivity (or corresponding current density),

$$\left.\begin{array}{l} \nabla \times \mathbf{H} = \quad \varepsilon \partial \mathbf{E}/\partial t \\ \nabla \times \mathbf{E} = -\mu \partial \mathbf{H}/\partial t \\ \nabla \cdot \mathbf{H} = 0 \\ \nabla \cdot \mathbf{E} = 0 \end{array}\right\}. \tag{37}$$

Taking the curl of both sides of the cross products in Eq. 37 gives

$$\left.\begin{array}{l} \nabla \times (\nabla \times \mathbf{H}) = \varepsilon \nabla \times \left(\partial \mathbf{E}/\partial t\right) \\ \nabla \times (\nabla \times \mathbf{E}) = -\mu \nabla \times \left(\partial \mathbf{H}/\partial t\right) \end{array}\right\} \tag{38}$$

or

$$\left.\begin{array}{l} \nabla \times (\nabla \times \mathbf{H}) = \varepsilon \partial \left(\nabla \times \mathbf{E}\right)/\partial t \\ \nabla \times (\nabla \times \mathbf{E}) = -\mu \partial \left(\nabla \times \mathbf{H}\right)/\partial t \end{array}\right\}. \tag{39}$$

Substituting from Eq. 37,

$$\left.\begin{array}{l} \nabla \times (\nabla \times \mathbf{H}) = -\mu \varepsilon \partial^2 \mathbf{H}/\partial t^2 \\ \nabla \times (\nabla \times \mathbf{E}) - \mu \varepsilon \partial^2 \mathbf{E}/\partial t^2 \end{array}\right\} \tag{40}$$

utilizing a vector identity,

$$\nabla \times (\nabla \times \mathbf{V}) = \nabla (\nabla \cdot \mathbf{V}) = -\nabla^{\mathbf{2}} \mathbf{V}, \tag{41}$$

where $\mathbf{V}$ is a general vector, Eq. 40 can be expressed as

$$\left.\begin{array}{l} \nabla^2 \mathbf{H} = \mu \varepsilon \, \partial^2 \mathbf{H}/\partial t^2 \\ \nabla^2 \mathbf{E} = \mu \varepsilon \, \partial^2 \mathbf{E}/\partial t^2 \end{array}\right\}. \tag{42}$$

Considering Eq. 36 for a conducting medium, Eq. 42 can be expanded:

$$\left.\begin{array}{l} \nabla^2 \mathbf{H} = \sigma \mu \partial \mathbf{H}/\partial t + \mu \varepsilon \, \partial^2 \mathbf{H}/\partial t \\ \nabla^2 \mathbf{E} = \sigma \mu \partial \mathbf{E}/\partial t + \mu \varepsilon \, \partial^2 \mathbf{E}/\partial t \end{array}\right\}. \tag{43}$$

The solution of Eq. 42 is a traveling wave of $\mathbf{H}$ or $\mathbf{E}$, respectively. Although these are separate wave equations, they represent physically inseparable quantities. Consequently, an electromagnetic wave can be ideally represented by plane waves as shown in Fig. 5. If the electric field equation in Eq. 42 is expanded, then

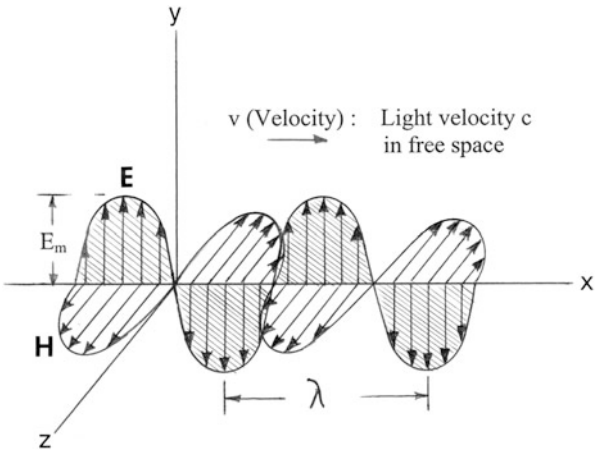


Fig. 5 Electromagnetic plane wave. Electric (E) and magnetic field vectors are shown in a plane parallel to the y-z plane for the wave traveling from *left* to *right* along the x-axis

$$\left.\begin{aligned}\frac{\partial^2 E_x}{\partial x^2}+\frac{\partial^2 E_x}{\partial y^2}+\frac{\partial^2 E_x}{\partial z^2}&=\mu\varepsilon\frac{\partial^2 E_x}{\partial t^2}\\ \frac{\partial^2 E_y}{\partial x^2}+\frac{\partial^2 E_y}{\partial y^2}+\frac{\partial^2 E_y}{\partial z^2}&=\mu\varepsilon\frac{\partial^2 E_y}{\partial t^2}\\ \frac{\partial^2 E_z}{\partial x^2}+\frac{\partial^2 E_z}{\partial y^2}+\frac{\partial^2 E_z}{\partial z^2}&=\mu\varepsilon\frac{\partial^2 E_z}{\partial t^2}\end{aligned}\right\}. \tag{44}$$

Considering the case represented by Fig. 5, $E_x = E_z = 0$; E_y is a function of x but not of y or z, and Eq. 44 reduces to

$$\frac{\partial^2 Ey}{\partial x^2}=\mu\varepsilon\frac{\partial^2 Ey}{\partial t^2}. \tag{45}$$

In its simplest form, Eq. 45 has a solution

$$E_y = f\,(x-vt), \text{or } f_1(x-vt)+f_2(x+vt) \tag{46}$$

where $v = 1/\sqrt{\mu\varepsilon}$ and represents the wave phase velocity in the medium characterized by permeability, μ, and permittivity. The function f (x−vt) represents any function of the quantity (x−vt), including a sine wave represented in Fig. 4.

It is of interest to note that in free space, the wave velocity will be light velocity, c, where

$$c = 1/\sqrt{\mu_o\varepsilon_o}. \tag{47}$$

A sinusoidal solution of Eq. 45 would then be

$$E_y = \sin(x-ct). \tag{48}$$

Correspondingly, solving for H_z in a formalism similar to Eq. 44,

$$H_z = \sin(x - ct), \tag{49}$$

and Eqs. 48 and 49 characterize the electromagnetic plane wave in Fig. 5 in free space.

On examining the wave formalism of Fig. 5, it can also be assumed that a wave in which the electric field is described by

$$E_x = 0; E_y = E_m \cos\beta(x - vt); E_z = 0. \tag{50}$$

The coefficients E_m and β are often referred to as the wave amplitude and phase constant, respectively, where β determines the frequency of the sinusoidal variation:

$$\beta v = \omega, \tag{51}$$

and Eq. 50 becomes

$$E_x = 0; E_y = E_m \cos(\omega t - \beta x); E_z = 0 \tag{52}$$

where $\omega = 2\pi f$ (f = frequency in Hz (Hertz or s^{-1}). Correspondingly, the wavelength, λ, in Fig. 5 is of course given by

$$\lambda = v/f = 2\pi/\beta; v = \omega/\beta = 2\pi f/\beta \tag{53}$$

The second Maxwell equation in Eq. 36 can be written as

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E} \tag{54}$$

as it applies to Fig. 5 and find that Eq. 50 can be expressed by

$$\begin{aligned} &\partial B_x/\partial t = 0; \partial B_y/\partial t = 0; \partial B_z/\partial t = -\partial E_y/\partial x \\ &= -\beta E_m \sin(\omega t - \beta x) \end{aligned} \tag{55}$$

where by integration with respect to time it is found that

$$B_x = 0; B_y = 0; B_z = (\beta/\omega)\, E_m \cos(\omega t - \beta x) = (\beta/\omega) E_y \tag{56}$$

where from Eq. 53,

$$E_y = vB_z. \tag{57}$$

Equation 57 as it applies to Fig. 5 describes a simple plane wave traveling in a dielectric medium where E_y and B_z are identical in form and perpendicular to one another and the direction of travel along the x-axis in Fig. 5. Since the wave velocity as noted previously is

$$v = 1/\sqrt{\mu\varepsilon} \tag{58}$$

and from Eq. 10 $B = \mu H$, we can write Eq. 57 as

$$E_y = v\mu H_z = \left(\mu/\sqrt{\mu\varepsilon}\right)H_z = \sqrt{\frac{\mu}{\varepsilon}}H_z \tag{59}$$

where the ratio of E to H is

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} \tag{60}$$

it is historically referred to as the intrinsic (wave) impedance in a medium. Note that in free space as a medium

$$v = c = 1/\sqrt{\mu_o \varepsilon_o} \tag{61}$$

and η in Eq. 60 becomes $\sqrt{\mu_o \varepsilon_o} = 4\pi c \times 10^{-7}$ ohms where values are substituted for μ_o and $\in_o$ as given previously. This is of course consistent with E (or E_y) in volts per meter and H (or H_z) in amperes per meter. Substituting for the velocity of light ($c \cong 3 \times 10^8$ m/s) results in a value η in Eq. 60 of 120π or 377 Ω for electromagnetic (light) wave propagation in free space.

It can be observed from Eq. 45 that the electromagnetic wave equation in free space ($\mu = \mu_o$, $\varepsilon = \varepsilon_o$; $\mu_o \varepsilon_o = 1/c^2$) can be rewritten as

$$\frac{\partial^2 E_y}{\partial x^2} - \left(\frac{1}{c^2}\right)\frac{\partial^2 E_y}{\partial t^2} = 0 \tag{62}$$

with reference to Fig. 5 or more generally as

$$\frac{\partial^2 E_m}{\partial x^2} - \left(\frac{1}{c^2}\right)\frac{\partial^2 E_m}{\partial t^2} = 0, \tag{63}$$

where, as noted, E_m is the wave amplitude. Similarly, for H_z in Fig. 5,

$$\frac{\partial^2 H_m}{\partial x^2} - \left(\frac{1}{c^2}\right)\frac{\partial^2 H_m}{\partial t^2} = 0. \tag{64}$$

Electromagnetic Wave Reflection, Refraction, and Transmission at the Free Space (Air)/Matter Interface

As implicit from Fig. 5, an electromagnetic wave traveling through space has half its energy vested in the electric field and half in the magnetic field. If the wave enters some different medium, there will be a new distribution of this energy. This assumes the new medium is a dielectric material, a magnetic material, or a

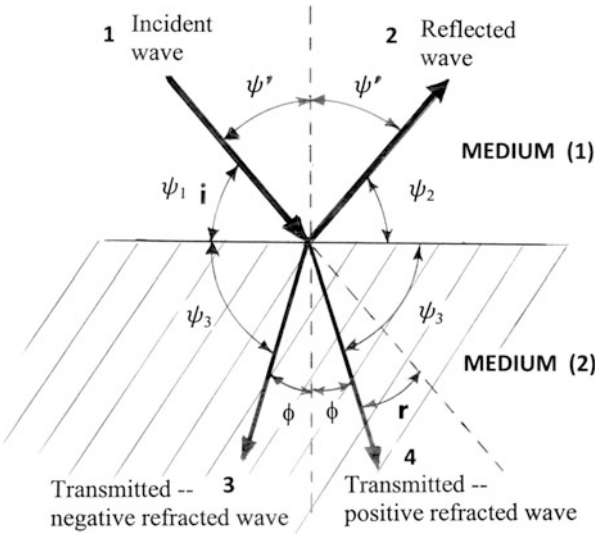


Fig. 6 Electromagnetic waves at boundary surfaces

conducting material. Figure 6 shows schematically all possible wave conditions at an interface or boundary separating medium (1) from medium (2).

Since in Fig. 6 the transmitted (refracted) wave travels more slowly than the incident and reflected waves, the wavelength is less, and it can be observed that

$$\frac{\lambda_3}{\cos X_3} = \frac{\lambda_1}{\cos X_1} = \frac{\lambda_2}{\cos X_2}. \tag{65}$$

However, since the incident and reflected waves travel in the same medium, $\lambda_1 = \lambda_2$ and $X_1 = X_2$; the incidence angle equals the reflection angle. The refraction (positive in Fig. 6) is characterized by

$$\frac{\cos X_1}{\cos X_3} = \frac{\lambda_1}{\lambda_3} = \frac{V_1}{V_3} = \frac{v(\text{medium}(1))}{v(\text{medium}(2))}. \tag{66}$$

For the case where medium (1) in Fig. 6 is air where $v = c$, it can be written generally as

$$\frac{\cos X_1}{\cos X_3} = \left(\frac{c}{v}\right) = n \tag{67}$$

where n is the index of refraction. Recall that

$$\left.\begin{aligned} c &= 1/\sqrt{\mu_o \varepsilon_o} \\ v &= 1/\sqrt{\mu\varepsilon} \end{aligned}\right\}, \tag{68}$$

the index of refraction becomes

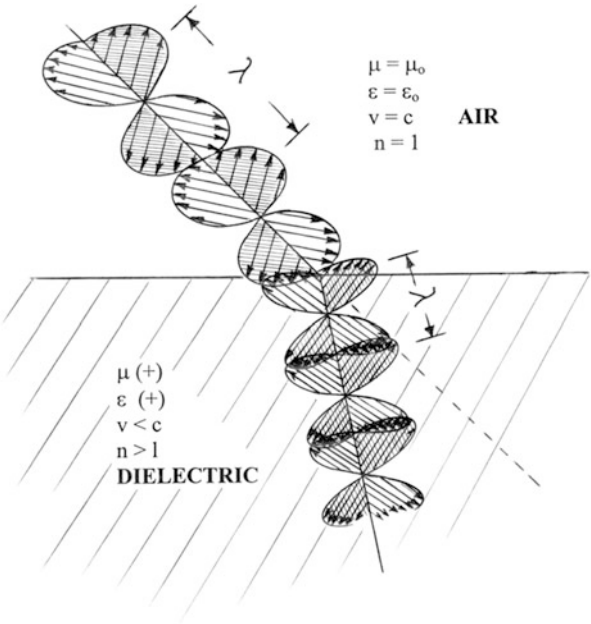


Fig. 7 Electromagnetic wave refraction in a dielectric material

$$n = \sqrt{\mu\varepsilon/\mu_o\varepsilon_o}. \tag{69}$$

For a dielectric medium (or material) as medium (2) in Fig. 6, $\mu = \mu_o$ and

$$\cos X_1/\cos X_3 = n = \sqrt{\varepsilon/\varepsilon_o} = \sin i/\sin r = \sin\psi'/\sin\phi \tag{70}$$

which is the more usual form of Snell's law (named after Dutch astronomer Willebrord Snellius (1580–1626)). Figure 7 illustrates the electromagnetic wave behavior traveling from air through a dielectric material with permittivity ε. Equations 67 and 70 indicate that the wave velocity relative to air is reduced in the dielectric by an amount v/c. This reduction is expressed by n for ε and μ positive. For metals such as good conductors Ag or Au, there is only reflection except when medium (2) in Fig. 6 is a very thin film. This is because metals have negative permittivity (−ε) at visible wavelengths (~400–800 nm). Similarly, medium (2) can have negative permeability (−μ). A material with either ε or μ negative is opaque to electromagnetic radiation (light). Ideally, thin metal films which allow transmission dependent upon the film thickness can be designed to vary the reflected energy versus the transmitted energy, and such devices are referred to as an optical beam splitter, allowing a portion of the incident beam to be reflected and a remaining portion transmitted.

Figure 8 shows an example of image refraction for a black straw in a beaker of water. Note this is image refraction and not wave refraction shown in Fig. 6. Moreover, the geometry of the photograph is not ideal for measuring true dihedral

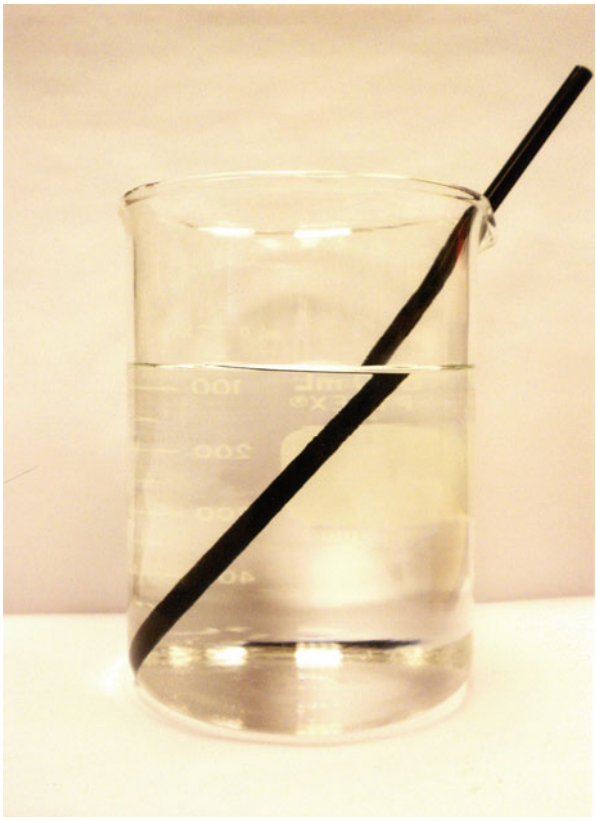


Fig. 8 Optical image refraction

angles. The index of refraction for water is 1.33, so the wave velocity in the water would be $3/1.33 \times 10^8$ m/s. Table 1 illustrates the positive index of refraction for a number of transparent, dielectric materials. Commonly, dielectric materials are characterized by a relative dielectric constant (K_r):

$$K_r = (\varepsilon/\varepsilon_o) = n^2. \tag{71}$$

It should also be noted in Table 1 that the values are approximate because they do not account for a small variation of index (n) with wavelength which is called dispersion. Dispersion can be visualized when a beam of light is split into different (rainbow) colors by a prism. This causes refraction at different angles.

In discussing dielectric materials, it may be useful to refer to the capacitor illustrated in its simplest form in Fig. 9. Here, two conducting plates are separated by a distance s. In Fig. 9a, the plates, with an area A, are separated in air (as the dielectric where $K_r = 1$). In Fig. 9b, a dielectric material has been inserted. The capacitance is given by

Table 1 Refractive index and dielectric constants for selected transparent materials

Material	n[a]	Kr (n^2)	ε (10^{-12} F/m)
Air	1	1	8.9
Water	1.33	1.77	15.8
Fluorite (CaF_2)	1.43	2.04	18.2
Calcite ($CaCO_3$)	1.49	2.22	19.8
Crown glass	1.52	2.31	20.6
Sodium chloride (NaCl)	1.54	2.37	21.1
Quartz (SiO_2)	1.54	2.37	21.1
Emerald ($Be_3Al_2(SiO_3)_6$)	1.57	2.46	21.9
Ruby/sapphire (Al_2O_3)	1.77	3.13	27.9
Cubic zirconia (ZrO_2)	2.17	4.71	41.9
Diamond	2.42	5.86	51.2
Silicon[b]	4.24	17.98	160.0

[a]Usually measured at $\lambda = 589$ nm (yellow double + sodium line)
[b]Not optically transparent unless thin film

$$C = \frac{A\varepsilon}{s} (\text{in Farads/m}) \tag{72}$$

where A is the plate area and s is their separation. The capacitance for the air dielectric (Fig. 9a) versus the dielectric material (Fig. 9b) is the ratio

$$C/C_o = \varepsilon/\varepsilon_o = K_r. \tag{73}$$

It can be observed in Table 1 that the permittivity or relative dielectric constant can provide the positive (intrinsic) index of refraction, or vice versa (Eq. 71). It might also be recalled that the first capacitor was the Leyden jar which could be air filled or filled with water to increase the permittivity (and capacitance) as shown in Table 1.

Wave Dispersion: Phase and Group Velocity and the Electromagnetic Photon

As shown in Fig. 7, an electromagnetic wave entering a dielectric or other medium causes the wavelength to shorten and the velocity of the wave phase to decrease. While it is useful to illustrate this concept using simple wave structures, the specific nature of electromagnetic wave interaction with matter, especially at the atomic (or electronic) level, is not tractable. Albert Einstein was one of the first to observe this dilemma and insisted that light impinging upon a photoconducting material would more practically consider a wave packet having energy characterized by M. Planck:

$$E = h\nu = hc/\lambda \tag{74}$$

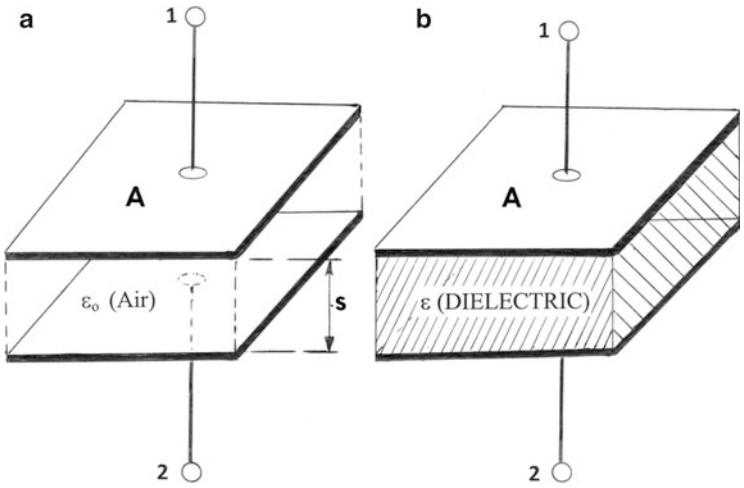


Fig. 9 Simple, parallel plate capacitor. (**a**) Air dielectric ($\varepsilon = \varepsilon_o$) and (**b**) dielectric material (ε). Voltage (potential difference V) applied between 1 and 2 to produce an electric field intensity E = V/s

where $\nu = f$ (frequency in Hz or s^{-1}).

These wave packets, called photons, would also be characterized by a group velocity, and dispersion can be represented by such a wave packet in the context of the model shown in Fig. 10. In this model, the phase velocity for the group of waves or wave packet interior is

$$v = v_p = \omega/\beta, \tag{75}$$

where $\omega = 2\pi f$ and β is the phase constant, while the speed of the group or group velocity is

$$v_g = d\omega/d\beta. \tag{76}$$

A common situation in free space is expressed by

$$v_p v_g = c^2 = 1/\mu_o \varepsilon_o. \tag{77}$$

Figure 9 depicts what occurs when a stone is thrown into a pool of water as the waves disperse from the point of origin at $t = 0$. In this case, the phase velocity of water waves is known to be

$$v_p = g/\omega = g/2\pi f, \tag{78}$$

where g is the gravitational constant and f is the wave frequency (or ν). Since $\beta = \omega^2/g$, differentiation gives

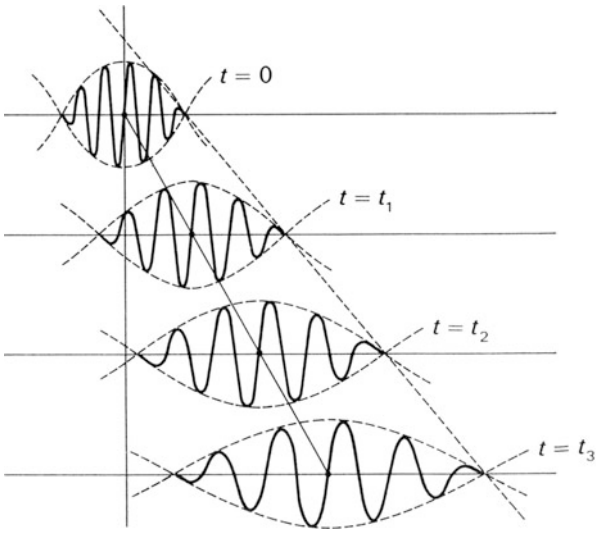


Fig. 10 Dispersion of a wave packet. Visualize a stone thrown into a pool of water where surrounding waves disperse from this point of origin for the disturbance

$$v_g = d\beta/d\omega = 2\,\omega/g = v_p/2. \tag{79}$$

In the case of angular dispersion of electromagnetic waves, n varies with λ:

$$dn/d\lambda < 0, \tag{80}$$

and the corresponding group velocity is given by

$$V_g = c\left(n - \lambda\frac{dn}{d\lambda}\right)^{-1} \tag{81}$$

for homogeneous media. Since $v_p = c/n$,

$$V_g = V_p - \lambda\frac{dv_p}{d\lambda} \tag{82}$$

from Eq. 81.

In 1922, A. Compton showed that photons (or wave packets) could be described by momentum (p) proportional to their frequency (ν) or

$$p = E/c = h\nu/c = h/\lambda. \tag{83}$$

This is the so-called de Broglie relationship (from the 1923 dissertation by L. de Broglie):

$$\lambda = h/p \tag{84}$$

which characterizes the wave-particle relationship in a generalized way, also termed wave-particle dualism. As applied to optical photons, mass estimates have been around 10^{-69} kg. This can be compared to the measured mass for the electron as 9×10^{-31} kg or the proton mass of 1.67×10^{-27} kg, a difference of roughly 10^{-38} kg when comparing a photon wave packet with an electron wave packet! However, in practice, the optical or electromagnetic photon is assumed to have zero rest mass.

References

Agassi J (2007) Science and its history: a reassessment of the historiography of science, Boston studies in the philosophy of science. Springer, New York

Moliton A (2007) Basic electromagnetism and materials. Springer, New York

Electromagnetic Color and Color in Materials

Contents

Abstract

It can be noted in ▶ Fig. 9 of chapter "Examples of Materials Science and Engineering in Antiquity" that color, represented by dyes and pigments (representing organic and inorganic materials, respectively), represents a sweeping application of materials fundamentals throughout human history. Even before Mayan blue (▶ Fig. 9a of chapter "Examples of Materials Science and Engineering in Antiquity"), Egyptian blue and Chinese blue were discovered and developed as will be described in chapter "▶ Serendipitous Nanotechnology in Antiquity." In this chapter, color in glass (an amorphous structure) is discussed in contrast to color in crystalline materials. Color in crystals often indicates some features of electrical behavior: conductors versus semiconductors for transparent crystals. Structural color arising from the diffraction or interference of electromagnetic waves in the visible portion of the spectrum is described in the context of the so-called photonic crystals, providing a contrast between dye and pigment augmented absorption and electronic absorption by substitutional impurities in crystal structures, including the creation of color centers such as F-centers where an electron occupies a lattice site normally occupied by a negative ion or anion. Finally, the concept of negative index of refraction is briefly discussed as this relates to the so-called stealth materials or materials structures related to photonics or photonic crystal structures, the idea that creating intricate combinations of negative permeability and negative permittivity materials systems can exhibit invisibility.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_4

Introduction

In some respects, color is a matter of perception when we consider the ability of a person to see color. People can be "trained" to see or identify specific colors, but this may be at variance with a person who has some form of "color blindness." However, color in the visible portion of the electromagnetic wave spectrum can be related to systematic or selective absorption phenomena, and this can be quantified by absorption spectroscopy. In a pragmatic sense, color and other aspects of visibility or invisibility become extinct or described in a different context for a sightless person.

It might be of interest to note that Maxwell made the first color photograph in 1861. This utilized the three primary colors: red (R), green (G), and blue (B). Superimposed, these three colors produce white light since they cover the visible portion of the spectrum R (750 nm) → B (400 nm). This primary color concept formed the basis for early television where three close-packed color phosphors created color in a similar fashion, along with the evolution of flat screen displays which employ the same primary color concepts. These will be explored in more detail where device concepts are presented in later chapters.

Color in nature, including mineral crystals, will be presented in this chapter along with other color mechanisms including optical (or photon) dispersion, interference, and diffraction. Examples include insect color and iridescence observed in butterflies and beetles, photonic crystals illustrated by close-packed (FCC) silica (SiO_2) nanometric-size spheres in opal, creating color zones where the sphere sizes change from ~400 nm (blue) to 750 nm (red). Color arising from electrons occupying negative ion crystal lattice sites (forming an F-center) will be illustrated in this chapter. Consequently, fundamental structural or microstructural issues as these relate to or control color in materials will be described. In many materials, color changes represent fundamental microstructure changes which also influence properties. For example, in some materials, particularly crystalline materials, a color change will be accompanied by a change in electrical conductivity. That is, a material may be a good conductor of electricity corresponding to one color, while at another color it may be a semiconductor or even become electrically insulating. These phenomena are related to different chemical or ionic species occupying lattice sites or other structural or microstructural domains or sites giving rise to optical (wave) absorption phenomena which in turn produce color.

In addition, the concept and reality of negative index of refraction ($-n$) in terms of engineered materials or metamaterials, including photonic crystals, will be discussed. This involves materials or material systems which exhibit negative permeability ($-\mu$) and permittivity ($-\varepsilon$) simultaneously. Under these conditions, electromagnetic waves are reflected or refracted backward, and a regime exhibiting negative index will appear invisible in the range of wavelengths involved, which can include microwaves, white light, infrared (IR), etc. These cloaking materials involve optical band width concepts similar to electronic bandwidth phenomena and can have profound effects as innovative materials, particularly optical devices.

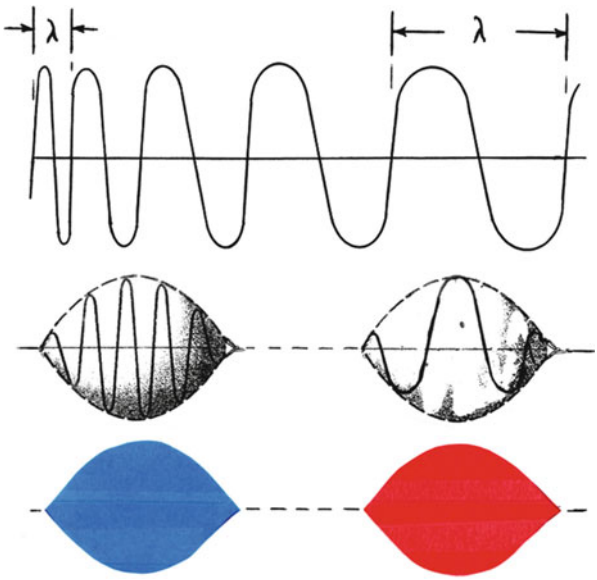


Fig. 1 Visible electromagnetic spectrum concept from blue to red ($\Delta\lambda \cong 400$–750 nm) color components

Light, or that portion of the electromagnetic spectrum we see, is composed of a range of wavelengths ($\Delta\lambda$) ranging from red (700–800 nm) to blue or violet (350–450 nm) color portions. Either the electric or magnetic plane wave component (▶ Fig. 5 of chapter "Electromagnetic Fundamentals") could be represented as illustrated in Fig. 1. However a light beam might be thought of as a collection of or stream of photons having characteristic energies or a range of energies $\Delta\varepsilon = hc/\Delta\lambda$. Ideally, color is what is perceived or measured in the transfer of this information from the optic nerve to the brain. Indeed, because of the energies associated with color, the skin can sometimes detect physiological effects which can also be differentiated as colors.

In the dispersion of a wave packet as shown in ▶ Fig. 10 of chapter "Electromagnetic Fundamentals," the wavelength within the packet actually increases as the envelop is stretched, and ideally the energy of the packet is decreasing. As an optical photon, the visible spectrum color would be changing from red to blue or in that direction. This dispersion or spatial separation of white light is commonly observed as material dispersion or chromatic dispersion in a prism where white light refracts at different angles (angular dispersion) causing it to split into the color components represented by wavelength regimes, $\Delta\lambda$, exhibiting a wavelength-dependent refractive index illustrated by ▶ Eq. 78 of chapter "Electromagnetic Fundamentals." This is illustrated in Fig. 2a, while Fig. 2b shows the same effect for a rainbow where raindrops essentially replace the prism in Fig. 2a. It is notable in looking at Fig. 2b that the observer must be facing away from the light source (the sun). Also, as illustrated in Fig. 2a, each color component ($\Delta\lambda$) is associated with a corresponding photon characteristic of that color and its corresponding energy or range of energies, ΔE. Table 1 compares the energy ranges for different photon types, including the visible portion of the electromagnetic spectrum.

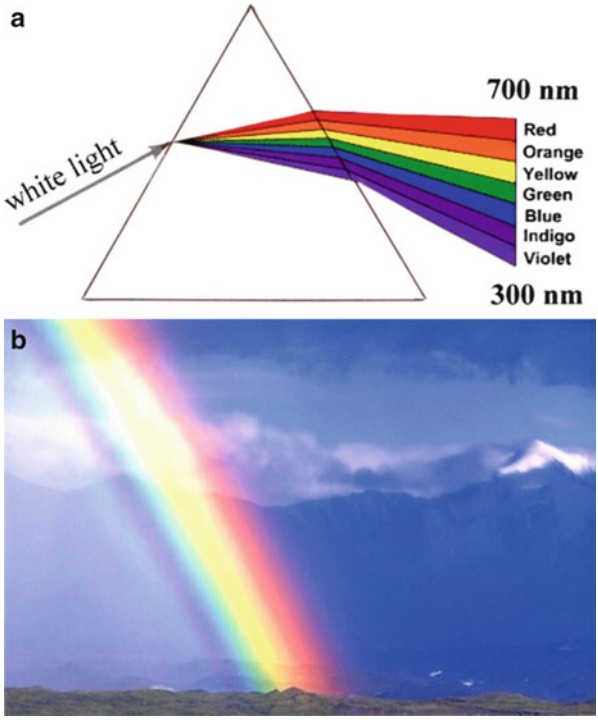


Fig. 2 Chromatic dispersion examples. (**a**) Material dispersion in a glass prism causing the spatial separation of white light into its component (photon) colors and corresponding wavelength ranges. (**b**) Raindrop-induced optical dispersion creating a rainbow. Note the similarity in color sequencing in (**a**) and (**b**)

Color in Glass

Historically, natural (volcanic) glass or obsidian was actually used in the Stone Age to fashion tools and weapons, and, as noted in chapter "▶ A Brief History of Metals," when people learned to smelt ores for metal production, the rock mass or gangue produced a slag which rose to the surface of the molten metal as a pre-glass vitreous material. The Egyptians are credited with developing this slag process into glass making using sand and lime, and glass-making technology grew rapidly during the late Bronze Age (▶ Fig. 1 of chapter "A Brief History of Metals"). As other ingredients were added to the glass-making process, colors were noted and in time artisans produced unique, colored glass objects using secret formulations and process variations (Shackelford and Doremus 2006).

The color of glass invariably depends on the absorption (or emission) of light or its component photons. This is characterized by the interaction of photons with the structure of glass which is a silicate (Si-O) framework containing atoms or ions (charged atoms) either within the framework or Si-O network, or substituting for the Si in its atomic site, or both. In addition colloidal metals or metal oxides can also aggregate within the network structure and create selective absorption of photons. Substitutional impurities (atoms or ions) are referred to as *chromophores*, while those within the silicate network and colloidal aggregates are called *stuffed derivatives*. Figure 3 illustrates several examples of glass structures or silicate networks.

Table 1 Photon wavelengths and energies

Photon type	λ	Energy (hν) (eV)
Gamma ray	<0.1 nm	7×10^5
X-ray	0.1–10 nm	$120–1.2 \times 10^5$
UV	10–390 nm	3–124
Visible	350–800 nm	1.6–3.4
IR	800 nm–1 mm	1.24×10^{-3}–1.6
Microwave	1 mm–1 m	1.24×10^{-6}–1.24×10^{-3}
Radio	1 mm–10^5 km	12.4×10^{-9}–1.24×10^{-3}

eV electron volts

Aside from the fundamental Si-O unit $(SiO_3)_n$, the network itself is irregular and referred to as an amorphous structure in contrast to a regular structure or crystal structure. Ideally, these silicate networks and the impurity atoms or ions are charge balanced, but localized ion or atom configurations give rise to *color centers* which are responsible for selective absorption and production of color as it is seen or perceived. In some cases, the energetic photons (in the sunlight for example) can induce electron transfer or valence changes which can alter the color either permanently or momentarily. Permanent changes are seen in glass left in the sun for long periods of time. Color control or switching is illustrated in Fig. 3d which illustrated photochromism or the changing of glass from clear to "colored" or darkened in the sun changing to clear indoors. This is caused by an intervalence transfer similar to the production of reactive oxygen species for Cu and Ag ions in microbial killing as discussed in chapter "► A Brief History of Metals" (Part I "The Ages of Metals and Materials"). Color variations can also be caused by high-energy photon (Table 1) irradiation of glass creating defects which act as color centers to create selective visible photon absorption and emission.

As illustrated in Fig. 3d, colloidal aggregates of Cu and Ag (or ions) produce darkened glass in sunlight where photons of energy hν create neutral Ag (Ag°). When the photons are reduced or eliminated indoors, neutral Ag becomes ionized (Ag^+), and the copper valence goes from Cu^{2+} to Cu^+. This process of intervalence transfer or reduction-oxidation (redox) is referred to as photochromism.

It can be noted in Fig. 3c and d that the color producing component sits in silicate channels which form an irregular network in contrast to the indigo molecule in the regular (crystalline) palygorskite channels creating blue color as illustrated in ► Fig. 9c of "Examples of Materials Science and Engineering in Antiquity." Indeed, the insertion of a dye or pigment molecule (organic or inorganic) into an otherwise transparent material such as a clear polymer will also create selective photon absorption and create color. Even in opaque materials color can be perceived by photon reflection.

Color in Crystalline Materials

In contrast to the irregular, nonperiodic (amorphous) structures characteristic of glass as depicted in Fig. 3, crystalline materials are periodic, and regular lattice

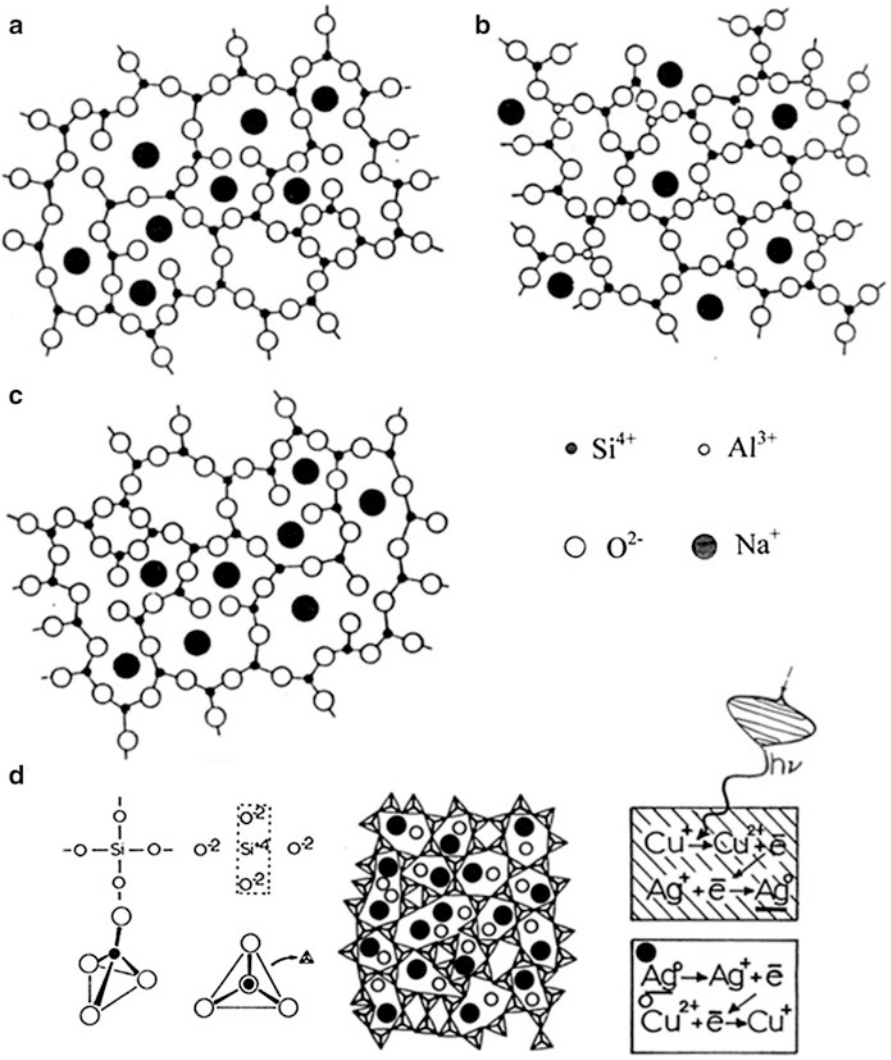


Fig. 3 Glass structure. (**a**) Soda-lime (clear) glass (75 % SiO_2 + Na_2O; CaO). (**b**) Alumina-silica (clear) glass (Al_2O_3 add to SiO_2). (**c**) Colored glass by impurity addition. (**d**) Photochromic glass. Intervalence transfer darkens in sunlight for photons of energy hν interacting with the colloidal aggregates or ionic substitutions of Cu and Ag. The silicate block is $SiO_4{}^{-4}$

arrays created by the spatial replication of unit cells which are variances of the 14 Bravais lattices representing seven crystal systems and are inducted by light energy. A similar process can be induced in thin films of WO_3 by an electric current (or voltage) where the valence change from W^{6+} to W^{5+} induces a color change. This process is referred to as electrochromism, and materials exhibiting this color change are electrochromic materials. In these structures, neutral atoms or ions

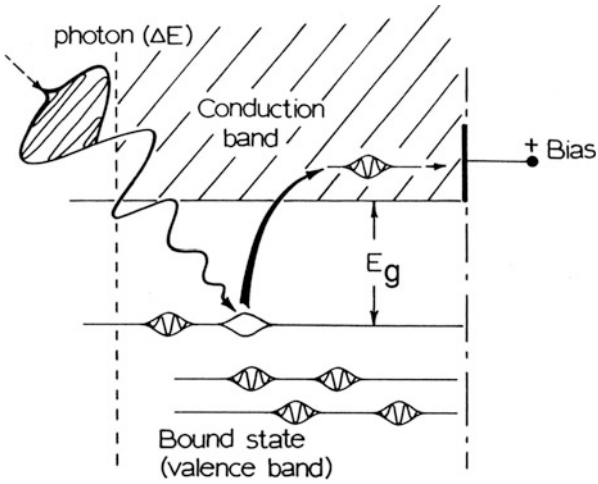


Fig. 4 Photon excitation of electron in the valence band to the conduction band for a band gap energy, E_g

occupy regular lattice sites, and deviations from this regularity are characterized by crystal defects. These include a variety of point defects: atoms or ions missing from regular lattice sites forming a *vacancy*, atoms or ions displaced from the regular lattice site and residing as an *interstitial*, and *vacancy-interstitial* pairs of positive (cation) and negative (anion) ions which in spite of their displacement provide charge balance in the structure. Cation variations or valence substitutions can also maintain charge balance such as substitution for C^{4+} in one of two cation vacancy sites to maintain charge balance in a structure of the form ($A^{2+}X^{2-}$), where A and X designate cation and anion species (such as $CaO:Ca^{2+}O^{2-}$). Similarly, substitution of A^{2+} and A^{4+} in two A^{3+} sites will maintain charge balance but create a local imbalance around the lattice site which can influence photon absorption. In addition, anion vacancies can also be occupied by electrons to maintain charge neutrality. These ionic arrays are referred to as *color centers* and act similar to color centers in glass in altering visible spectrum photon absorption and emission.

Precious gem stones such as diamond, sapphire, ruby, etc. epitomize practical, commercial examples of these phenomena along with related materials science and engineering fundamentals (Hall 2006). These fundamentals characterize photon absorption, ionization, and emission, where ionization utilizes the absorbed photon energy to transfer electrons in the valence band of a crystal to the conduction band. The band model for a semiconductor such as silicon is illustrated in Fig. 4. Here a photon having energy (ΔE) greater than the energy gap (E_g), which in silicon is 1.1 eV, forces an electron in the valence band into the conduction band, forming a so-called *hole* in the valence band (a vacant electron site). The signal detected will be proportional to the photon energy or their density (or intensity to number/unit area). This concept is the basis for photoconduction where a biased component in an electrical circuit may not allow for sufficient current flow until exposure to light creates larger current by creating more conduction band electrons. In the presence of an applied voltage (or electric field), the holes in the valence band can also move

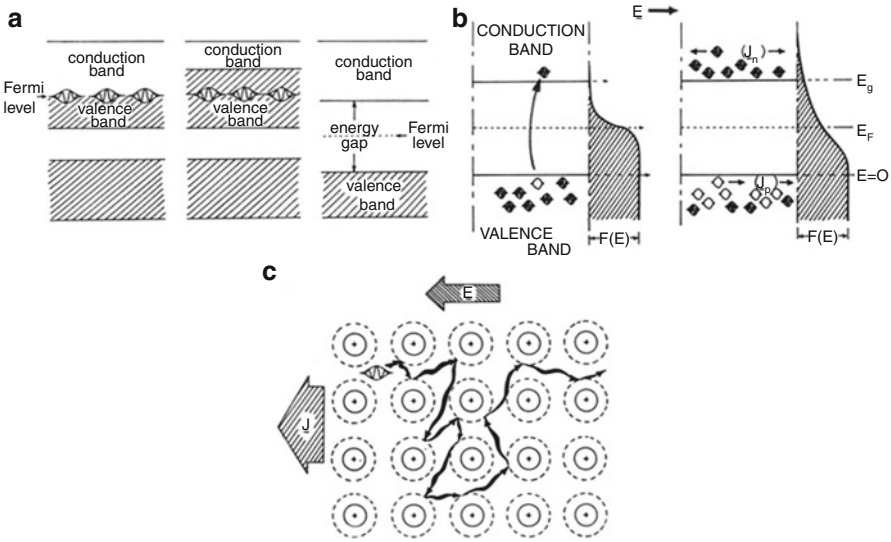


Fig. 5 Simple models for electrical conductors, semiconductors, and insulating materials. (**a**) Conductors represented by valence band overlap with the conduction band (two left models) and a band gap material (semiconductor for small gap and insulator for large gap). (**b**) Electron and hole production. (**c**) Normal electron conduction: interrupted drift opposite to the electric field direction

by allowing electrons to hop into that site, causing current flow opposite to the electron flow in either the valence or conduction band. Of course the application of an external voltage will, in the absence of electromagnetic radiation, move electrons across the energy band gap, E_g.

Figure 5 embellishes this concept for a band gap material in contrast to a metal or alloy considered to be a regular electrical conductor in contrast to a semiconductor. In Fig. 5a, ordinary conductors are illustrated by the two diagrams to the left. The far left diagram shows electrons at the Fermi level representing the top of the valence band overlapping the conduction band. The middle diagram shows a better conductor with electron energies above the Fermi level (E_F) in the overlapping valence band. The energy-band diagram at the right in Fig. 5a is similar to Fig. 4. Figure 5b shows the band structure diagram in Fig. 5a in the context of an electron energy distribution function (F(E)) which can be considered a depiction of thermal effects where more electrons are transferred to the conduction band from the valence band with increasing temperature. Figure 5c shows an electron moving through a crystal lattice under the action of an applied field, E, which produces a net current flow, J. The electron, acting as a negatively charged particle, will have its path interrupted by collisions with the atoms of the structure. These interruptions are averaged as the electron mean free path length, l, and conduction (or conductivity) is described in terms of this parameter as

$$\sigma = n_c e^2 l / m\bar{v} \tag{1}$$

where n_c is the number of electrons in the conduction band, e is the electron charge, m is the electron mass, and $\overline{v}$ is the net increase in the electron velocity or its drift velocity in the applied field.

In an intrinsic semiconductor such as pure silicon, the number of electrons per unit volume in the conduction band will be

$$N_n = 2\left(\frac{2\pi mkT}{h^2}\right)^{3/2} F(E), \tag{2}$$

where

$$F(E) = F(E_g) \cong e^{-(E_g - E_F)/kT}. \tag{3}$$

Here k is the Boltzmann constant, h is Planck's constant, and T is the temperature; $E_F = E_g/2$. If it is assumed that the effective mass for holes generated in the valence band in Fig. 5b is equal to the electron mass, we can write Eq. 2 as

$$N_n = N_p = 2\left(\frac{2\pi mkT}{h^2}\right)^{3/2} e^{-Eg/kT} \tag{4}$$

Correspondingly, the total conductivity in an intrinsic semiconductor becomes

$$\sigma = N_n e\mu_n + N_p e\mu_p \tag{5}$$

where μ_n and μ_p are the electron and hole mobilities, and mobility is defined by the applied field intensity and the drift velocity:

$$\mu = \overline{v}/|E| \tag{6}$$

Substituting for $N_n = N_p$ from Eq. 4 in Eq. 5 then

$$\sigma = 2e\left(\mu_n + \mu_p\right)\left(\frac{2\pi mkT}{h^2}\right)^{3/2} e^{-Eg/kT} \tag{7}$$

for conductivity in an intrinsic semiconductor.

It is apparent from the diagrams and equations that conductivity improves with decreasing energy gap (E_g) and increasing temperature. For very large energy gap materials, there is little or no conductivity and these materials are generally referred to as insulators. But as it turns out, the band gap itself can be modified when other atoms are substituted for the regular atoms, such as silicon in the example under discussion. Raising the temperature significantly can also increase conductivity by moving electrons from the valence band to the conduction band. This is true for diamond.

Figure 6 illustrates a comparison of band gaps for silicon and carbon (or diamond) with both having the same diamond cubic crystal structure or unit cell. In Fig. 6a, the Si band gap represents $E_g \cong 1.1$ eV while the C band gap represents (to scale) $E_g \cong 5.5$ eV. In contrast to Si, C (diamond) is a wide band gap material, normally an insulator. However, as shown in the diagrams of Fig. 6b, the band gap can be modified by substitution of an atomic species which contributes

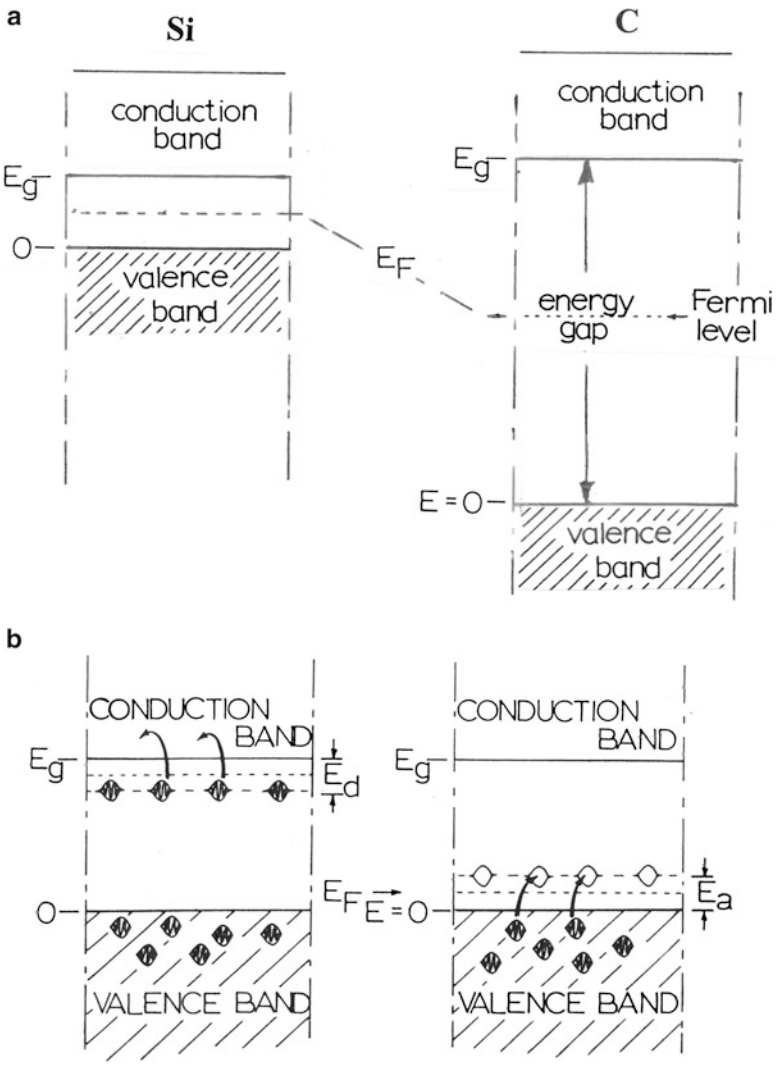


Fig. 6 Comparison of band gap principles. (**a**) Small gap (E_g = 1.1 eV) semiconductor (Si) compared with large gap (E_g = 5.5 eV) insulator (diamond, C). (**b**) Band gap modification with donor (*left*) and acceptor (*right*) substitutional atoms producing electron and hole substates, respectively

electrons to the structure (a donor atom creating an energy gap E_d below the conduction band) or conversely an atomic species which contributes holes to the structure (an acceptor atom creating an energy gap E_a above the valence band). Obviously, in the context of Fig. 4, the situations shown in Fig. 6b change the photon absorption characteristics and the corresponding color or color perception in the case of diamond (C) which in its pure form is clear.

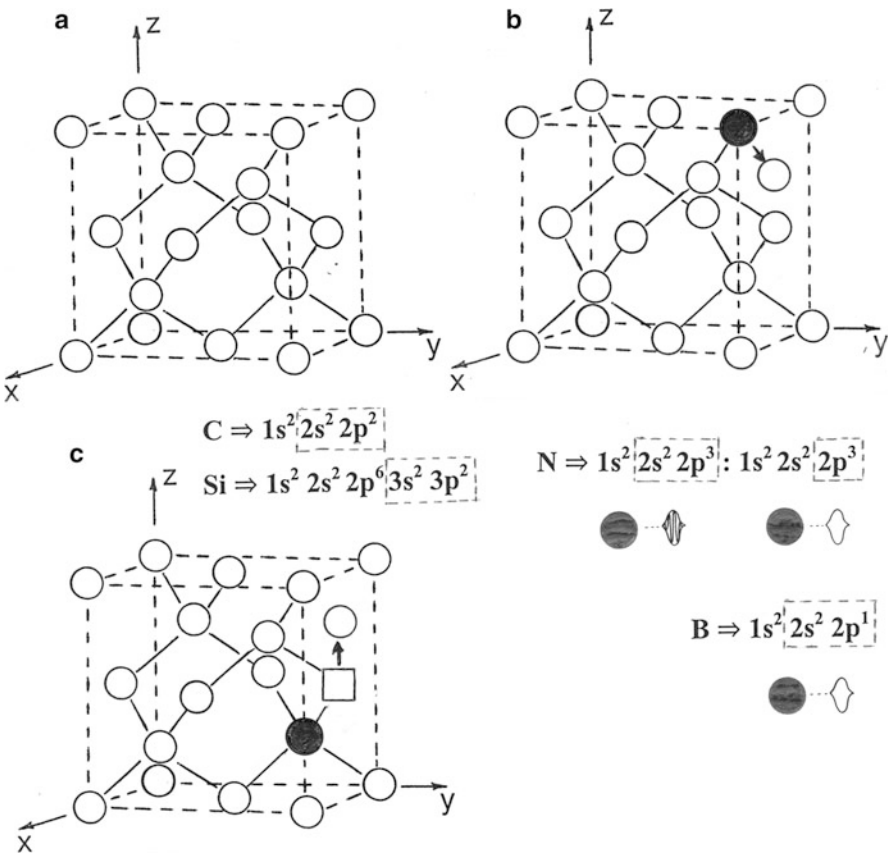


Fig. 7 Diamond cubic crystal unit cell models for diamond (C) and semiconducting silicon (Si) with corresponding electronic structure models. (**a**) Pure unit cell. (**b**) Substitutional atom-interstitial formation. (**c**) Substitutional atom-vacancy-interstitial complex. Donor and acceptor atoms are connected to an electron or hole

Figure 7 shows the diamond cubic crystal structure unit cell for both diamond (C) and Si. In the case of diamond (C), the atomic number or number of electrons is 6 while for Si it is 14. However, both C and Si are covalently bonded and share electrons as indicated in the dotted enclosure of their electron structure in Fig. 7a for second-shell (n = 2) electrons. Correspondingly, Si shares electrons to stabilize the third-shell (n = 3) electrons. Energetically stable atomic (or electronic) configurations are those where shells (quantum states) and quantum substates are filled.

Figure 7b shows an atom like nitrogen (atomic number, Z = 7; seven electrons in the neutral atom) substituting on a carbon atom site. The carbon atom has been displaced to form an interstitial atom (arrow). Because the electronic structure of nitrogen can either share all five of its L-shell (n = 2) electrons (5) or only the three in the p-substate, relative to the stable, covalent bond sharing to fill the p-substate, N^{5+} will have an extra electron C and donate that electron to the band gap where E_d

$\cong$ 1.7 eV in Fig. 6b. This can occur naturally in diamond or when diamonds are synthesized in the laboratory in high pressure-high temperature systems where nitrogen substitutes on C sites. One N atom in 10^5C atoms will create a yellow color which intensifies with nitrogen concentration. Correspondingly, when boron is substituted in the diamond lattice for C, the covalent (shared) bonds are satisfied, but there is a deficiency of one electron (an acceptor situation) represented by a hole as shown in Fig. 7. However, *B* has only a very small acceptor energy, $E_a \cong 0.05$ eV, and while it will alter the visible photon absorption to create a blue color, the alteration of the band structure in creating a semiconductor is considerably less.

Figure 7c shows a substitutional impurity and a vacancy-interstitial pair (arrow) in contrast to the substitutional impurity-interstitial pair in Fig. 7b. These atomic configurations are also augmented by excess electrons or deficient electrons (holes). In semiconductors like silicon, these substitutional elements are called dopants, and intrinsic semiconductors (like Si) doped with electron donor atoms produce a more negative (or n-type) semiconductor while when doped with electron acceptor atoms (creating holes just above the valence band) create a more positive (or p-type) semiconductor. The conductivity expressed by Eq. 7 is then referred to as extrinsic semiconduction since the external dopant provides some control over the conduction process, changing the concentration of electrons (for n-type) or holes (for p-type).

Since heating crystal structures to high temperatures can allow for atoms to be dislodged from their lattice sites to form vacancy-interstitial pairs, or even more complex atomic configurations if substitutional atoms are present, the color can often be altered by heating. In addition, even natural diamonds can be colored because natural processes can change the crystal structure. For example, diamonds formed near radioactive elements emitting alpha particles (α) can cause interstitial-vacancy pairs to be created since α-particle energies are around 5 MeV, more energetic than cosmic ray photons (Table 1). These particle clusters (composed of two protons and two neutrons: ideally He^{2+}) can knock atoms from their normal lattice sites. Similarly, irradiation of natural or synthetic diamonds by any energetic particle or radiation (proton or electron beams or X-ray photons, etc.) can create or alter diamond color. This is true for other crystalline materials as well.

Figure 8 illustrates the absorption spectra for a number of common gem crystals. In the case of ruby, the crystal basis is corundum or trigonal (hexagonal) alumina (Al_2O_3) which, in contrast to diamond, is primarily ionic binding: ${Al_2}^{3+}\,{O_3}^{-2}$. Small additions of Cr^{3+} to aluminum sites where the Cr^{3+} ions are larger than the Al^{3+} ions cause the clear corundum to color pink, creating a pink sapphire in contrast to the clear sapphire which is pure Al_2O_3. Higher concentrations of Cr^{3+} produce ruby as the pink color becomes red. Substituting Fe^{2+} and Ti^{4+} for two Al^{3+} ions in this structure will maintain charge balance but alter visible light photon absorption producing blue sapphire, while Fe^{3+} substitutional atoms produce yellow sapphire. These atomic configurations will yield a specific absorption spectrum which can be used to identify the corresponding defect configurations as illustrated generally in Fig. 8.

Figure 9 illustrates a variety of colors produced in diamonds by the creation of various defect configurations. In addition, Fig. 9 illustrates the manipulation of absorption, reflection, and transmission of the visible spectrum by the geometry of

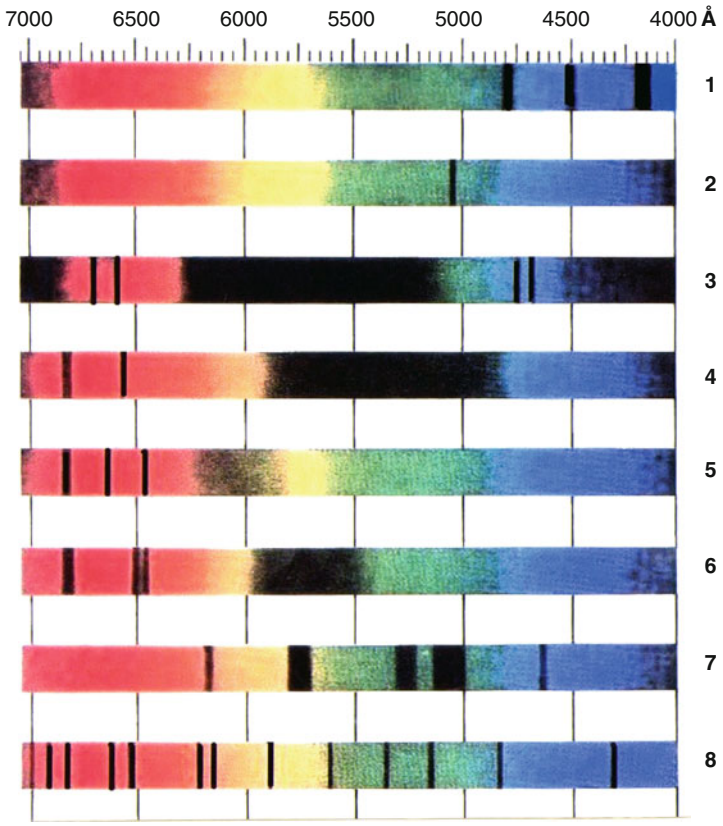


Fig. 8 Absorption spectra for some common single-crystal gem stones. *1* Pure natural diamond. *2* Natural diamond near radioactive source. *3* Ruby. *4* Red Spinel (72: 28 alumina/magnesia) often of the form $MgAl_2O_4$-Al_2O_3: MgO a cubic structure. *5* Emerald (or beryl: Be_3Al_2 $(SiO_3)_6$). *6* Alexandrite ($BeAl_2O_4$ with Cr^{3+} replacing some Al^{3+} ions). *7* Garnet (Fe_3Al_2 $(SiO_4)_3$). *8* Zircon (cubic zirconia (ZrO_2)) having the fluorite crystal structure

cutting of the gem stone. In the case of diamonds, they are marketed based on the nature of the "cut," the color, the weight (in carats: 1 carat weight is 0.2 g), and the cost (often referred to as the 4 Cs). Carat weight (CWT) is also referred to in fractions of one carat weight called points; one point equals 0.01 carat. Some popular cuts of diamond (especially the brilliant cut) also have a diameter which is related to the carat weight somewhat systematically. In contrast, it might be noted that Karat weight (K) of precious metals such as gold is based on 24 K = 100 %. Consequently 18 K Au = (18/24) × 100 = 75 %, etc.

It might be of interest to expand upon the concept of color centers described for glass coloring illustrated in Fig. 3 and color centers in ionic crystals in particular. Common color centers in simple ionic crystals such as NaCl or related AX ionic crystals involve the removal of the anion to create a vacancy which becomes occupied by an electron to maintain charge balance. This creates a unique absorption

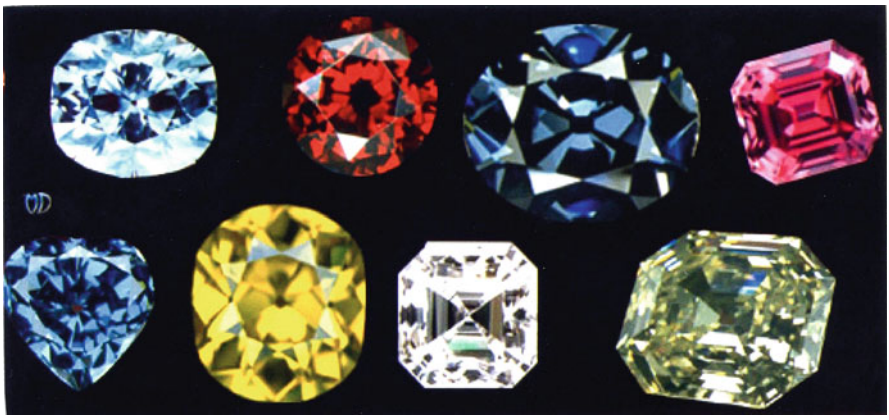


Fig. 9 Examples of diamond colors and gem stone cuts (facet geometries)

phenomenon which is illustrated in Fig. 10 for NaCl. Increasing concentrations of these electron charge-balanced anionic vacancies, called *F-centers* (after the German word, Farbe for color) create increasingly intense blue color. Similar F-center absorption is also created in ionic AX_2 crystals such as CaF_2 (Ca^{2+} F_2^-) as shown in Fig. 11. Figure 11b also shows a charged cation vacancy – anion interstitial pair which creates an absorption zone or color center that maintains charge balance. Figure 2d shows two F-center defects in CaF_2, while it might be noted that it could be possible to create large zones where all anions are removed, creating continuous F-centers (electron occupied anion vacancy sites) and leaving only the cations surrounded by electrons (shown as quantized wave packets in Figs. 10 and 11). This phenomena can occur in both NaCl crystal or CaF_2 crystal structures as shown in Figs. 10 and 11, respectively. Moreover, these electrons can combine with either the Na^+ cations in Fig. 10 or the Ca^{2+} cations in Fig. 11 to create neutral Na° or Ca° atoms or neutral metal zones (so-called solid-state metal colloids) which color these respectively crystals by selective white light absorption.

It might also be noted in Fig. 10 that the NaCl ionic structure is characterized by two interpenetrating face-centered cubic (FCC) crystal structures: Na-FCC/Cl-FCC. Correspondingly, Fig. 11 is a modification of the diamond cubic structure shown in Fig. 7a where four additional internal (8-cube) sites are occupied by anions (F^-) in Fig. 11a. These features are discussed in more detail in (Part IV "► Crystal Structures and Crystal Chemistry"), Chapter "► Crystallography Principles."

The complexity of color centers in ionic crystals such as NaCl or CaF_2 can be associated with other valence state cation substitutions which would necessitate cation vacancies or other adjustments to maintain charge balance. These defects coupled with F-centers would create different absorption arrays producing different colors. These features occur in natural CaF_2 (fluorite) which can have many colors, with F-centers induced by natural radiation (Fowler 1968; Schulman and Compton 1963). Figure 12 shows coloration in natural fluorite. The blue color in ► Fig. 22a of chapter "Examples of Materials Science and Engineering in Antiquity" occurs for

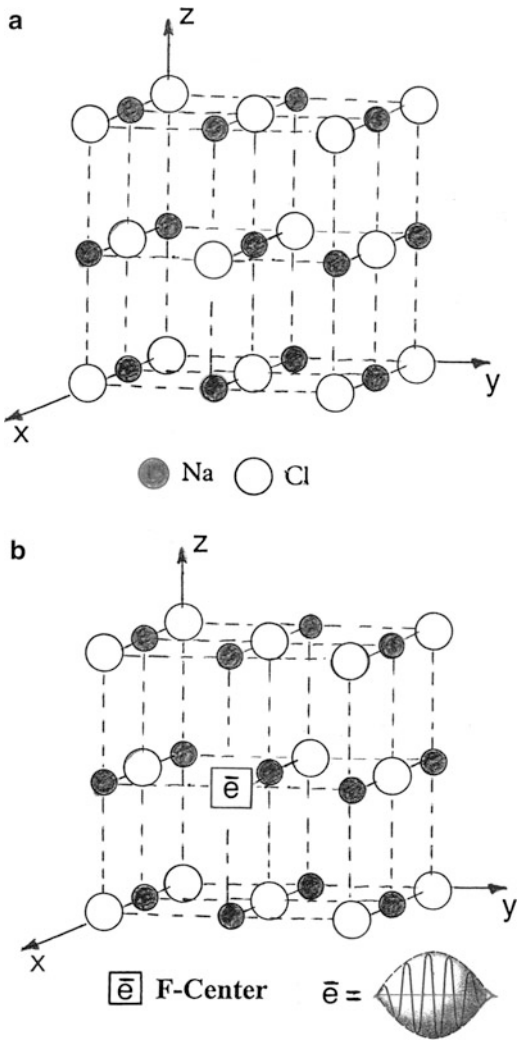


Fig. 10 F-center production in NaCl. (**a**) NaCl clinographic projection (or unit cell). (**b**) NaCl unit cell containing anion (Cl^-) vacancy occupied by an electron ($\bar{e}$) creating an F-center

F-centers, while the pink in ▶ Fig. 22b of chapter "Examples of Materials Science and Engineering in Antiquity" occurs for F-centers combined with Mn^{2+} cations substituting for the Ca^{2+} cations (Schulman and Compton 1962; Fowler 1968).

Structural Color and Photonic Crystals

Examples have been illustrated of color which originates by altering the selective absorption of white light by either the addition of pigments to materials or crystalline components (such as ▶ Fig. 9 of chapter "Examples of Materials Science and Engineering in Antiquity") or *chromophores* or color centers in materials or crystals

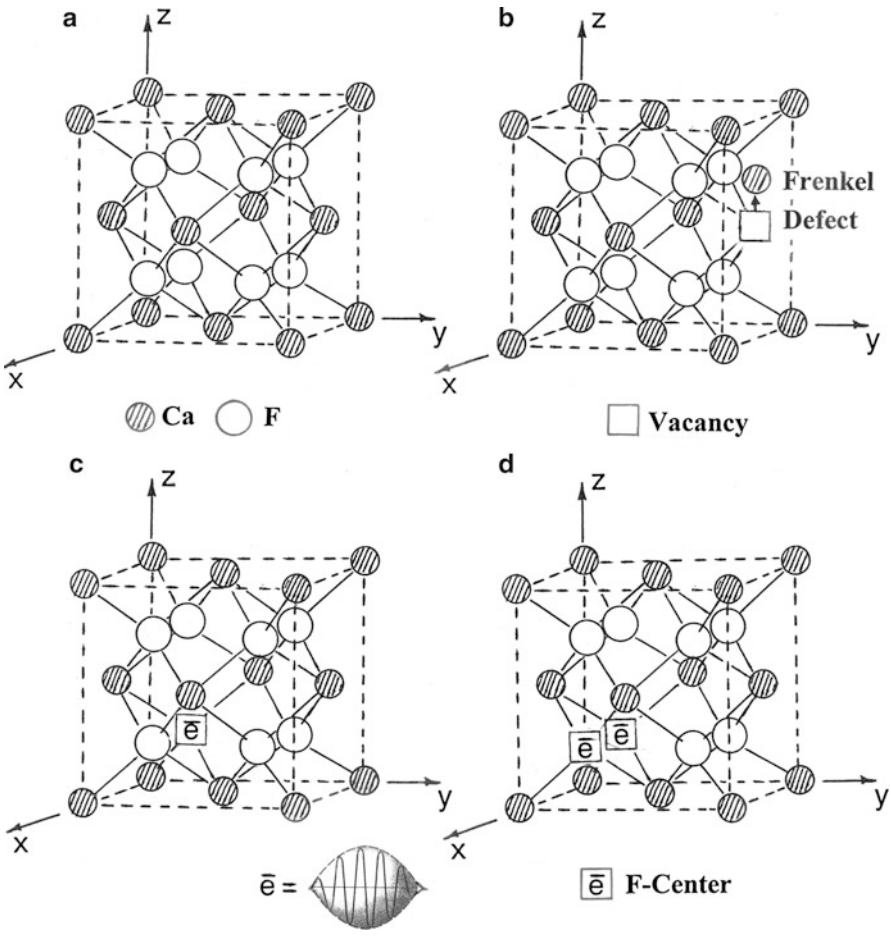


Fig. 11 F-center production in CaF_2. (**a**) CaF_2 clinographic projection (or unit cell). (**b**) Frenkel defect creation by shifting F^- (cation) to interstitial site. (**c**) CaF_2 unit cell containing anion (F^-) vacancy occupied by an electron ($\bar{e}$) creating an F-center. (**d**) Double F-center

(Figs. 3, 8, 9, 10, 11, and 12). However, color in nature often arises for purely physical or structural origin involved with diffraction or interference of light. Such color is often referred to as iridescence and readily recognized in the context of butterfly wings, beetle colors, and peacock feathers. Structural color also arises from systematic, even crystalline arrangements, of spheres or rods of selective permittivity or dielectric constants such as opal coloration. These are often referred to as *photonic crystals* (or optical metamaterials, also called *engineered optical structures*) (Joannopoulos et al. 2008). They can occur as 1-D, 2-D, and 3-D periodic structures.

Both thin-film *interference* and *diffraction* effects can be illustrated by a plane wave of light incident on a thin layer of thickness as shown in Fig. 13. Here, with

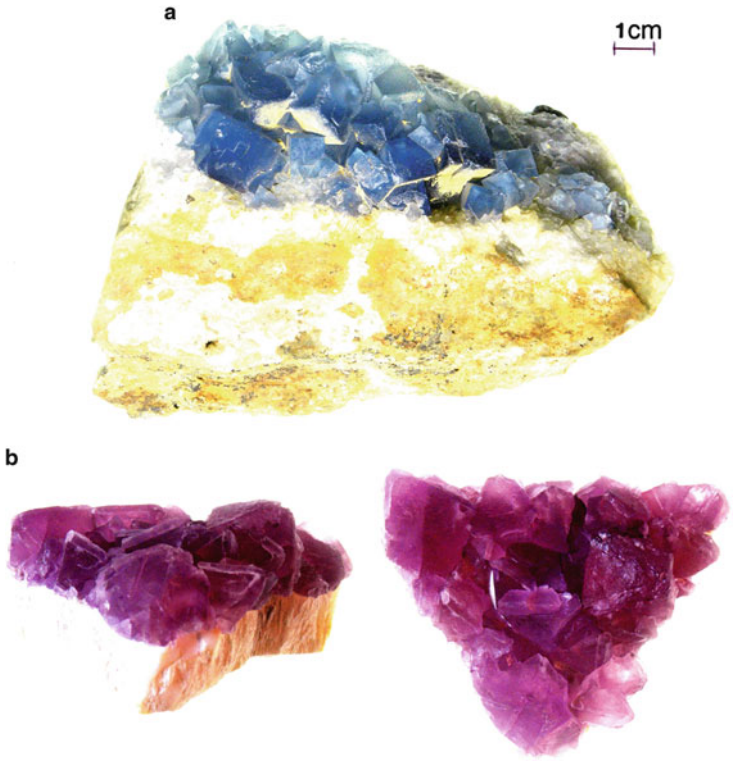


Fig. 12 Color of natural fluorite. (**a**) F-center blue coloration. (**b**) Pink coloration due to Mn^{2+} substitution and F-center complex

reference to the schematic of electromagnetic waves at boundary surfaces shown in ▶ Fig. 6 of chapter "Electromagnetic Fundamentals," the first-order wavelength observed will be given by

$$\lambda = 2ndcos\phi \tag{8}$$

for constructive interference for a transparent film or for opaque material,

$$\lambda = 2ndsin\psi' \tag{9}$$

for diffraction, with reference to Fig. 13a. These observed wavelengths (Eqs. 8 and 9) are also observed for 3-D metacrystals constructed from spheres of dielectric materials having permittivity, ε as shown in Fig. 12b. Examples of thin-film color development involve oil on pools of water, soap bubbles or foam on beer, and the like. These effects are also often dependent upon the viewing angle (or the viewer's angle), and this is particularly true for gratings or regular, spatial geometries where the effective dimension (or dimensions) will depend on the viewing angle or the incident wave angle (ψ' in ▶ Figs. 6 of chapter "Electromagnetic Fundamentals"

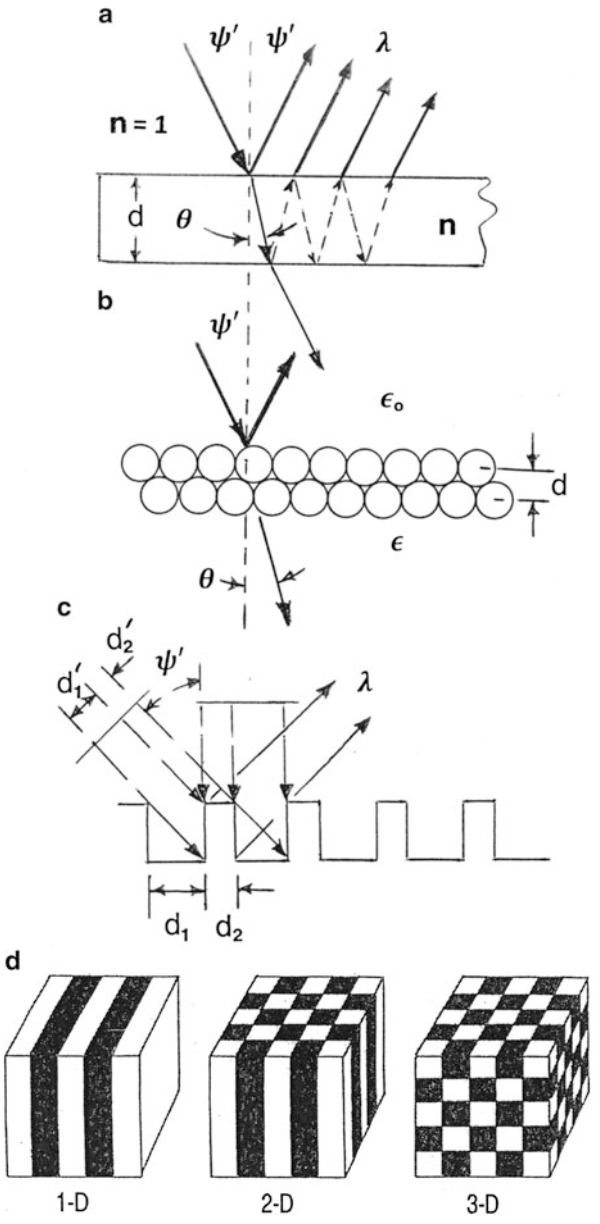


Fig. 13 Structural color fundamentals. (**a**) Interference and diffraction from thin films. (**b**) Diffraction from 3D dielectric photonic crystal. (**c**) Optical geometry for viewing angle color changes. d_1 and d_2 change with viewing angle ψ'. (**d**) 1-D, 2-D, and 3-D photonic structures

and ▶ 13a of chapter "Electromagnetic Color and Color in Materials"). This can be illustrated in the geometry shown in Fig. 13c. A common example of color produced by a 1-D diffraction grating having different (opponent) viewing angles occurs for a compact disc (CD). Figure 13d illustrates the three dimensions for photonic structures.

Figure 14 illustrates several examples of structural color. Figure 14a shows a CD which, because of two planes of curvature, shows color zones for a common viewing direction. Figure 14b shows a common butterfly. The wing section shown circled (Fig. 14b) is illustrated in the scanning electron microscope (SEM) image in Fig. 14c showing a tile-like scale structure. This structure is magnified in Fig. 14d which shows a 1-D, wide grating. This grating is observed to be further refined in the SEM image in Fig. 14e where the dimensions are consistent with the yellow portion of the spectrum: $\lambda \cong 550$–600 nm. In Fig. 14f color variations in a fire opal occur by variations in the diameters of the silica (SiO_2) spheres composing the opal as shown in the SEM image of Fig. 14g. The dimensions of the silica spheres in Fig. 14g are consistent with the yellow-orange color region shown circled in Fig. 14f.

The general concept of photonics involves the emission, transmission, amplification, detection, and modulation of light (visible portion of the electromagnetic spectrum), although it can be extended to other wavelength ranges. Although photonic structures (and crystals which are 3-D photonic structures shown in Fig. 14d) are in homogeneous media with dimensions comparable to wavelengths of interest, photonic materials and photonic crystals can be engineered or artificial structures which can manipulate permittivity and permeability. In addition, photonic materials can influence photons in the same way electronic (semiconductors) materials can influence electrons. That is, photons, behaving as waves and propagating through these engineered structures, can be selectively blocked, creating photonic band gaps characteristic of semiconductor band gaps illustrated in Fig. 6. Ideally, wavelength dependence as illustrated in Eqs. 8 and 9 show that even 1-D structures (Fig. 13d) can separate polychromatic (white) light into constituent, wavelength components-creating optical dispersion like a prism as shown in Fig. 2a. At other wavelengths, photonic materials can manipulate electromagnetic waves in novel ways beyond the capabilities of lenses, mirrors, and prisms.

Electromagnetic Metamaterials: Negative Index of Refraction

The ability to select structural components having specific permittivity (and corresponding index of refraction) to create photonic crystals or other, dimensional, engineered materials with dimensions that are fractions of the electromagnetic wavelengths. Metals, such as Au and Ag, have negative permittivity ($-\varepsilon$) at visible wavelengths, and materials with either ε or μ negative are, as noted previously, opaque to visible light. Negative μ has been observed in some ferromagnetic materials for microwaves. There are, however, no natural materials with both ε and μ negative.

In 1967, Victor Veselago (1968) theorized artificial or metamaterials having both negative ε and μ, exhibiting a negative index of refraction which would cause incident light or any other characteristic electromagnetic wave to be reflected (or refracted) back (reversed in direction) as illustrated in ▶ Fig. 6 of chapter "Electromagnetic Fundamentals." The choice of a negative sign for both ε and μ

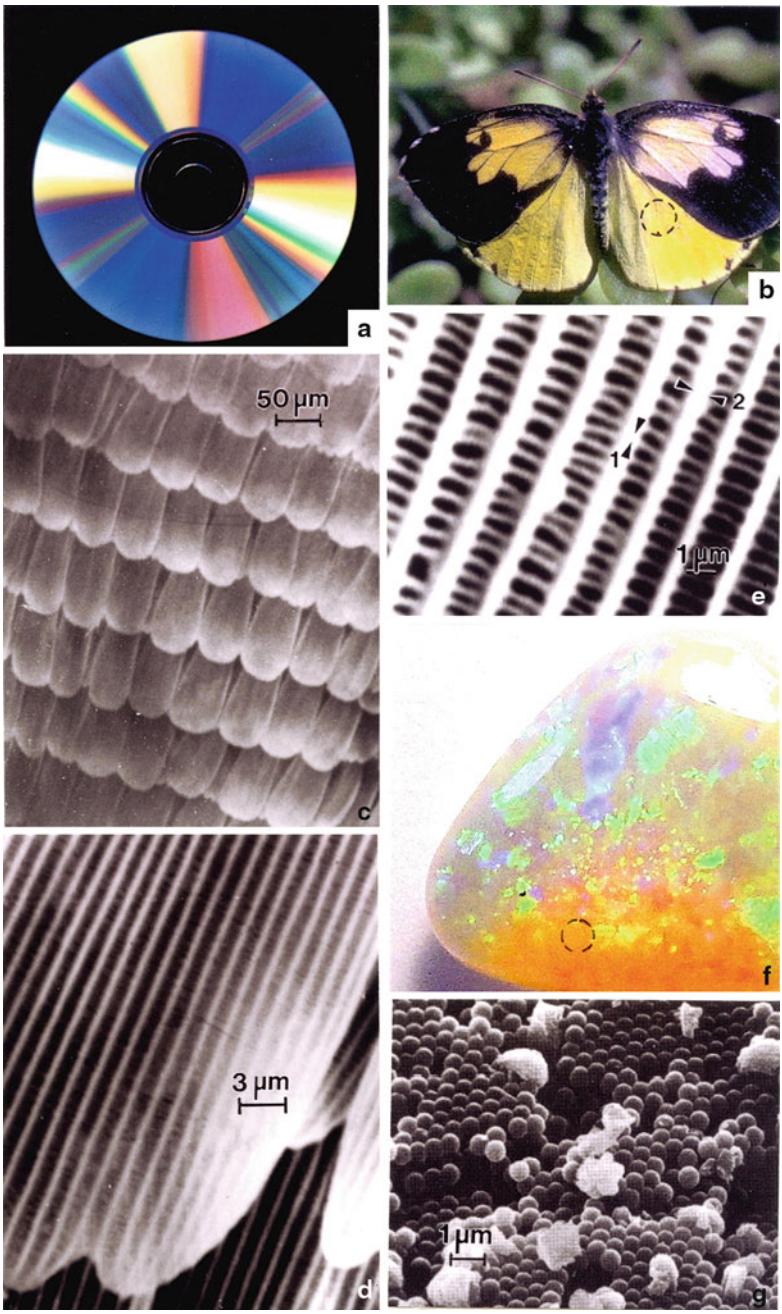


Fig. 14 Examples of structural color. (**a**) CD: compact disc. (**b**) Butterfly. (**c**) Butterfly wing scales. (**d**) and (**e**) Butterfly wing scale detail. (**f**) Fire opal color zones. (**g**) Silica sphere structure corresponding to circled area in (**f**)

does not cause any mathematical contradictions in the index of refraction expressed by ▶ Eq. 62 of chapter "Electromagnetic Fundamentals," but the appearance of a negative factor of refraction requires rewriting ▶ Eq. 62 of chapter "Electromagnetic Fundamentals" in a more general form.

$$n = \pm\sqrt{\mu\varepsilon/\mu_o\varepsilon_o}, \tag{10}$$

where the positive sign refers to ▶ Eq. 62 of chapter "Electromagnetic Fundamentals" and the negative sign is used when both μ and ε are negative.

The associated wave reversal for *negative index of refraction materials* (NIRM) is also characterized by antiparallel phase velocity. Such materials are also referred to as *left-handed materials* (Krowne and Zhang 2010). Negative index materials (NIM) would appear to be tractable if metals and specific ferromagnetic materials were combined in a novel photonic composite with the -ε and -μ components arranged in ordered, continuing (periodic) lattices or lattice arrays with dimensions that are fractions of the electromagnetic wavelengths of interest as illustrated schematically in Fig. 13d (Krowne and Zhang 2010). The first composite metamaterial was composed of Cu split-ring resonators in a periodic array of Cu wire conducting elements on interlocking sheets of fiberglass (Shelby et al. 2001). This NIM composite array transmitted microwave radiation at frequencies of 4–7 gigahertz ($\Delta\lambda = 1$ mm to 1 m). While photonic crystals can exhibit unusual electromagnetic wave propagation behavior such as phase and group velocity reversal, negative refraction (−n) has not been demonstrated in these systems. While NIMs have been developed for single visible light frequencies (Krowne and Zhang 2010; Turpin et al. 2014), there are no pass band arrays that operate on visible light. The achievement of such metamaterial arrays could allow for the development of coating-like arrangements which could render the coated regime invisible to an observer in ordinary light or within wavelength regimes (the IR, e.g., $\Delta\lambda = 900$–1,200 nm). This could be accomplished by growing nanowire arrays in insulator (dielectric) matrices or more complex, periodic nanocomposite structures on flexible substrates using 3-D printing or additive manufacturing processes (Berman 1971).

References

Berman M (1971) Regular-faced convex polyhedra. J Franklin Inst 291(5):329–352
Fowler WB (1968) Physics of color centers. Academic, New York
Hall J (2006) The encyclopedia of crystals. Hamlyn/Octopus Publishing Group Ltd, London
Joannopoulos JD, Johnson SG, Wina JN, Meade RD (2008) Photonic crystals: molding the flow of light, 2nd edn. Princeton University Press, New Jersey
Krowne CW, Zhang Y (eds) (2010) Physics of negative refraction and negative index materials. New York, Springer
Schulman JH, Compton WD (1963) Color centers in solids. Pergamon Press, Oxford
Shackelford JF, Doremus RH (2006) Aromic and glass materials: structure properties and processing. Springer, New York

Shelby RA, Smith DR, Shultz S, Nemat-Nasser SC (2001) Microwave transmission through a two-dimensional, isotropic, left-handed metamaterial. Appl Phys Letters 78(4):489–492

Turpin JP, Bossard JA, Morgan KL, Werner DH, Werner PJ (2014) Reconfigurable and tunable metamaterials: review of the theory and applications. Int J Antennas & Propagation 2014: article ID 429837, 18 pages http://dx.doi.org/10.1155/2014/429837

Veselago VG (1968) The electrodynamics of substances with simultaneously negative values of ε and μ. Sov Phys Usp 10(4):509–514

Part III

Evolution of Quantum Mechanics: Applications to Electrons in Atoms and Solid Structures

A Brief Introduction to Quantum Mechanics

Contents

Abstract

Utilizing the wave-optics equation derived by Maxwell, E. Schröinger inserted the de Broglie relationship to derive a time-independent equation with an embedded wave-particle dualism and energy quantization. This chapter illustrates the Schrödinger equation and its application to electrons composing atoms, an equation of motion for an electron. Simple illustrations examine electron energy quantization in atoms as well as so-called "free" electrons in a volumetric confinement such as an electrical conductor. This brief introduction to quantum mechanical principles as these relate to atomic structure represented by electrons occupying quantized energy states serves as a precursor to understanding matter or materials at the atomic or ionic level.

Introduction

It might be recalled that classical mechanics fundamentals are illustrated by Newton's laws of motion, namely, that a body at rest or in motion remains in that state unless acted upon by an outside force ($F = 0$), and

$$F = ma = \frac{d(mv)}{dt} = \frac{dp}{dt} \tag{1}$$

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_5

where m is the mass, a is the acceleration, v is the linear velocity, and p is the corresponding linear momentum. The total energy of this body (E) can be expressed:

$$E = (\text{kinetic energy}) + (\text{potential energy}) = mv^2/2 + U \tag{2}$$

Rewriting the momentum we can observe that

$$p = mv = 2\,(E - U)/v \tag{3}$$

or

$$p^2 = (mv)^2 = 2\,m(E - U) \tag{4}$$

where the relativistic mass, m, is given by

$$m = \frac{m_o}{\sqrt{1 - \left(\frac{v}{c}\right)^2}}, \tag{5}$$

and m_o is the rest mass.

Following M. Planck's quantization of light, A. Einstein interpreted Planck's equation (▶ Eq. 72 of chapter "Electromagnetic Fundamentals") as "particles" of light or photons. As noted in ▶ Eq. 81 of chapter "Electromagnetic Fundamentals," A. Compton showed that photons could be described by their momentum being proportional to their frequency at about the same time Louis de Broglie proposed that all "particles" of matter in the classical sense possessed a wavelike nature (or a wavelength, λ) and rewrote Compton's equation (▶ Eq. 81 of chapter "Electromagnetic Fundamentals") in the form expressed in ▶ Eq. 82 of chapter "Electromagnetic Fundamentals," now known as the de Broglie relationship. More importantly, de Broglie extended the quantized photon concept to electrons or "electron waves" in the context of matter waves. L. de Broglie in fact received the 1929 Nobel Prize in physics for "his discovery of the wave nature of electrons." This was 2 years after A. Compton received the Nobel Prize in physics.

Of course some of these concepts evolved from experimental evidence such as the work of Max von Laue around 1912 that showed that X-rays could be diffracted by crystals, thereby establishing their wave nature. Correspondingly, in 1914, von Laue received the Nobel Prize in physics "for his discovery of the diffraction of X-rays by crystals." Almost simultaneously, W. L. Bragg, at the age of 22, demonstrated the geometry of diffraction by the now well-known Bragg's law:

$$n\lambda = 2d_{hkl} \sin\theta, \tag{6}$$

where λ is the X-ray wavelength, d_{hkl} is the spacing between atomic planes in a crystal, and θ is the so-called Bragg angle or diffraction angle. Figure 1 illustrates this diffraction geometry (or the diffracted wave, **K**) referenced to the crystal surface plane and the incident X-ray beam vector, $\mathbf{K}_o$. W. L. Bragg was at the

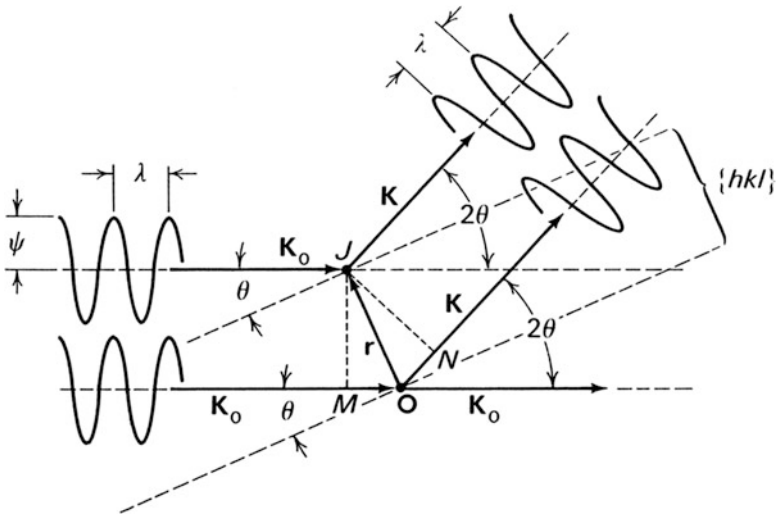


Fig. 1 Diffraction of plane waves having a wavelength, λ, from crystal planes separated, d_{hkl}. The wave amplitude is denoted, ψ. **r** is a vector connecting a reference atom in a crystal unit cell to any atom at position J. The path difference $\Delta = (OM + ON) = \lambda$

time a protege of his father, W. H. Bragg, a well-regarded crystallographer. Three years later, in 1915, they shared the Nobel Prize in physics "for their services in the analysis of crystal structure by means of X-rays." W. L. Bragg, at age 25, was the youngest to date (2013) to receive a Nobel Prize.

About a decade later, in 1927, C. J. Davisson and L. H. Germer showed that electrons were also diffracted from a nickel single crystal in the same manner as X-rays, and G. P. Thomson and A. Reid almost simultaneously showed that electrons also diffracted in an identical manner when transmitted through a thin nickel crystal, corresponding to the geometry of diffraction in Fig. 1. The Davisson-Germer experiment showed that electrons, particles with measurable mass, m, acted like waves, and confirmed the work of de Broglie (▶ Eq. 82 of chapter "Electromagnetic Fundamentals").

C. J. Davisson and G. P. Thomson shared the 1937 Nobel Prize in physics for their work.

Evolution of the Schrödinger Equation

During the period around 1924, Peter Debye proposed that if particles behaved as waves as argued by de Broglie (▶ Eq. 82 of chapter "Electromagnetic Fundamentals"), they should satisfy some sort of wave equation. This prompted Erwin Schrödinger in late 1925 to derive a wave equation for time-independent systems such as the hydrogen atom (a one-electron atom) and demonstrate correct energy eigenvalues (energy quantization for an electron in a bound state) (Moore 1992). This derivation considered

the electromagnetic wave equations developed by Maxwell (▶ Eqs. 61 of chapter "Electromagnetic Fundamentals" and ▶ 62 chapter of "Electromagnetic Fundamentals") along with the de Broglie relationship (▶ Eq. 82 of chapter "Electro magnetic Fundamentals").

An example of this derivation begins by rewriting ▶ Eqs. 61 of chapter "Electromagnetic Fundamentals" and ▶ 62 of chapter "Electromagnetic Fundamen tals" in a more general wave equation for propagation in a medium having permittivity and permeability ε and μ and a corresponding index of refraction, n. Assuming $E_m = H_m = \psi$ (the wave amplitude), we can write

$$\nabla^2\psi - \left(\frac{n}{c}\right)^2 \frac{\partial^2\psi}{\partial t^2} = 0. \tag{7}$$

Since $n = c/v = c/\nu\lambda$ (where ν is the frequency, f), Eq. 7 can be rewritten

$$\nabla^2\psi - \frac{1}{(\nu\lambda)^2} \frac{\partial^2\psi}{\partial t^2} = 0. \tag{8}$$

The wave amplitude, ψ, can be expressed generally as

$$\psi = \psi(x, y, z, t) = \psi\ (x, y, z)e^{2\pi i\nu t} \tag{9}$$

and substituting into Eq. 8 results in

$$e^{2\pi i\nu t}\nabla^2\psi + \frac{e^{2\pi i\nu t}(4\pi^2\nu^2)}{(\nu\lambda)^2}\psi = 0 \tag{10}$$

or

$$\nabla^2\psi + \frac{4\pi^2\psi}{\chi^2} = 0 \tag{11}$$

substituting for λ using the de Broglie relationship in ▶ Eq. 82 of chapter "Electromagnetic Fundamentals" and the momentum squared as expressed in Eq. 4, results in

$$\nabla^2\psi + \frac{8\pi^2 m(E - U)}{\hbar^2}\psi = 0 \tag{12}$$

If we let $\hbar = h/2\pi$, Eq. 12 above becomes

$$\nabla^2\psi + \frac{2m\,(E - U)}{\hbar^2}\psi, \tag{13}$$

which represents the time-independent form of the Schrödinger equation. As applied to the electron, Eq. 13 represents the electron wave-particle dualism since ψ relates to the wave amplitude and is referred to as the wave function, and m is the electron mass. Schrödinger actually shared the 1933 Nobel Prize in physics with P. A. M. Dirac "for the discovery of new particle forms of atomic theory."

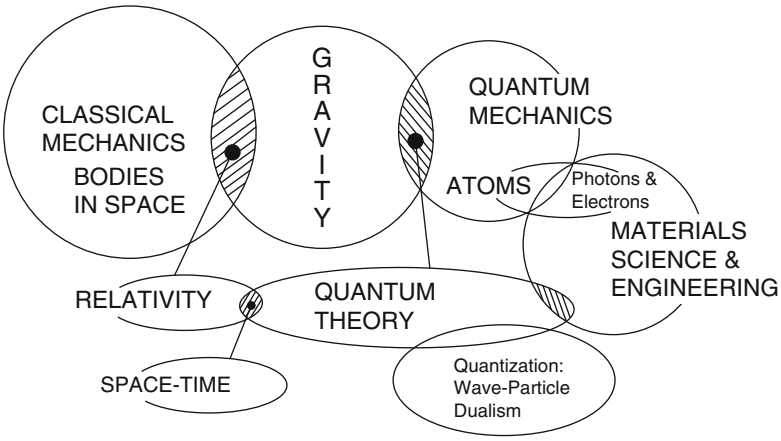


Fig. 2 Classical to quantum mechanical connections in space-time

Equation 13 is the basis for wave-particle dualism and energy quantization. In effect, it is an equation of motion for the electron. When $U = 0$, solutions for E in Eq. 13 are quantized, as will be shown later, and this forms the basis of quantum mechanics in contrast to classical mechanics. However, it is of interest to notice that inserting the de Broglie relationship into the electromagnetic wave equation, the classical expression for linear momentum (Eq. 3) was also utilized. Consequently, there is an intrinsic connection between classical and quantum mechanics implicit in deriving the Schrödinger equation (Eq. 13). Indeed, as implicit in Eq. 13, there is a connection to relativity, especially for the electron. Figure 2 provides a graphic representation of the connections between classical and quantum mechanics, including gravity. Recall that gravity is large for large bodies in the universe and correspondingly very weak for atoms and electrons. Space-time is also connected; however, without consciousness space and time become nonexistent. The connections illustrated in Fig. 2 also imply equations since everything has equations: the universe and everything in it, including materials and their constituent – atoms, electrons, and subatomic components.

Equation 13 is easily adapted to describing an electron, say in a cathode ray tube which may be accelerated from the filament by an applied anode voltage, V, where $V = U$, the potential energy. A much simpler case involves the electron in a solid conductor-like gold. We refer to electrons not bound to an atom of Au to be "free" electrons, and these are electrons in the overlapping valence and conduction band as depicted in ▶ Fig. 5a of chapter "Electromagnetic Color and Color in Materials." We can illustrate this electron within some bound state which is often depicted as an "electron in a box" (or a finite volume of a conductor such as Au) as shown in the sketch of Fig. 3. We assume, with reference to Eq. 13, that inside the box, $U = 0$ and that in order to assume it cannot escape the box, $U = \infty$ outside. Consequently, if the box has dimensions L, the solution of Eq. 13 as a second-order, linear, differential equation in the form

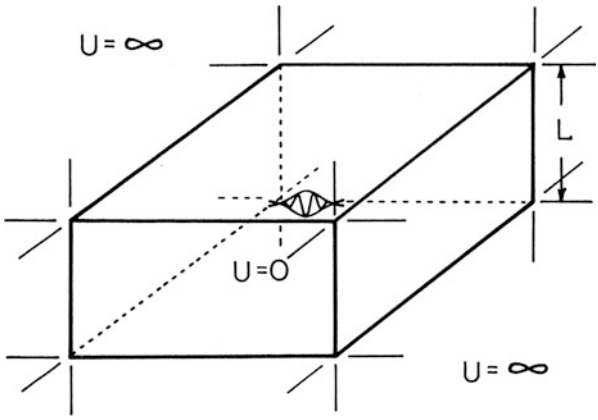


Fig. 3 Electron in a conducting box

$$\nabla^2\psi + \frac{2mE}{\hbar^2}\psi = 0 \tag{14}$$

becomes

$$E(x, y, z) = \frac{h^2}{8mL^2}\left(n_x^2 + n_y^2 + n_z^2\right) \tag{15}$$

$$\psi\,(x, y, z) = A\sin\pi n_x x \cdot \sin\pi n_y y \cdot \sin\pi n_z z \tag{16}$$

From Eqs. 15 and 16, we see that the electron energy, E (x, y, z), is quantized since $n_x, n_y, n_z = 1, 2, 3$, etc. and represent quantum states. The wave function ($\psi(x, y, z)$) for the electron as a plane wave is also similarly quantized. For a simple, one-dimensional linear (x-direction) reference for the electron in the box, we can write Eq. 15 as

$$E(x) = \frac{h^2 n_x^2}{8mL^2} \tag{17}$$

We can visualize atoms as electrons bound to a positively charged core (nucleus). In the classical mechanics sense, this is characteristically a mean circulating a planet as Earth and its moon. This could represent a single electron bound to a single proton. The potential energy, U, for this atomic configuration can be written as Coulomb's law:

$$U(r) = -\frac{Ze^2}{4\pi\varepsilon_o r} \tag{18}$$

where the nucleus charge is Ze, and for the hydrogen atom with only one proton and one "orbiting" electron (of charge $\bar{e}$), the value of Z, the atomic number (or number of protons or bound electrons) is equal to 1. The electron orbit has a radius of r, and ε_o is the permittivity of free space. Substituting Eq. 18 into Eq. 13 results in

$$\nabla^2\psi + \frac{2m}{\hbar^2}\left[E + \left(\frac{Ze^2}{4\pi\varepsilon_o r}\right)\right]\psi = 0 \tag{19}$$

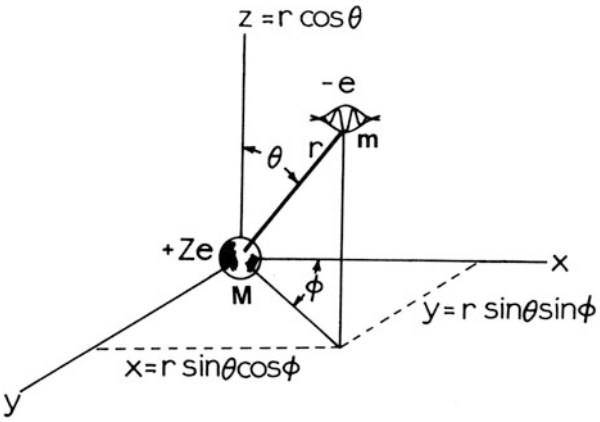


Fig. 4 Simple (geometry) rendering for the hydrogen atom showing Cartesian coordinates and spherical coordinate equivalents. e is the electron charge which is negative (−) for the electron and positive (+) for the proton (nucleus)

We can also include the nuclear mass, M, and replace m in Eq. 19 with mM/(m + M). However, considering the mass difference, m for the electron versus M for the proton initially for the hydrogen atom, the difference is inconsequential. To describe the actual position of the electron relative to the nucleus (or proton) in Cartesian, 3-D space requires a geometry depicted in Fig. 4 which superimposes spherical coordinates on the Cartesian coordinates for convenience: $(x, y, z) \Rightarrow (r, \theta, \phi)$.

A simple solution for the total energy, E, in Eq. 19 for the hydrogen atom depicted in Fig. 4 (for $Z = 1$) is

$$E = -\frac{me^4}{2(4\pi\varepsilon_o\hbar)^2}\left(\frac{1}{n^2}\right) \tag{20}$$

where n is the principal quantum number having integral values ($n = 1, 2, 3$, etc.) which quantize the atom's energy or the energy for the electrons which characterize or compose the atom. It is of course the electrons composing atoms which determine material behavior or the nature of solids, particularly solid crystals.

It is of interest to compare the implications of energy levels or quantum states of "free electrons" or those bound to a quantum well (or a "box") in the conduction band, with electron energy levels or quantum states composing or characterizing atoms. In the first case shown in Fig. 3, the energy is proportional to n^2, while for the second case shown in Fig. 4, the energy is proportional to $-1/n^2$. This circumstance is conceptually like turning the "box" in Fig. 3 inside out while maintaining the binding of the electron through a dimension $L = 2r$. Conceptualized in terms of energy levels (n), the two situations might be visualized as shown for comparison in Fig. 5. Here, the ground state, $n = 1$, is the lowest energy for the free electron in the box, or conducting volume, while for an atom, the electron in the ground state ($n = 1$) nearest to the atomic core or nucleus (composed of 1 proton and 1 neutron for each electron (Z)) also has the lowest energy (or the highest negative energy).

A more rigorous solution of Eq. 19 in spherical coordinates (Murr 1978; Muller-Kirsten 2012), where

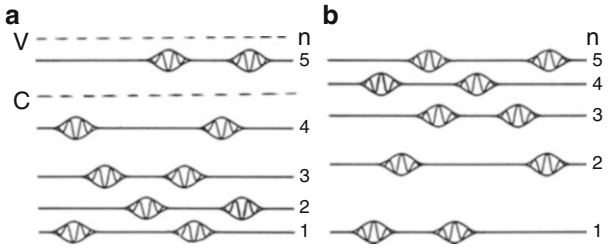


Fig. 5 Comparison of quantized electron energy states. (**a**) Electrons in a box (free electrons in a solid). V and C represent the top and bottom of the overlapping valence and conduction bands, respectively (▶ Fig. 5a of chapter "Electromagnetic Color and Color in Materials"). (**b**) Electron (s) in an atom

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin\theta}\frac{\partial^2}{\partial\phi^2} \tag{21}$$

even for the hydrogen atom, shows that energy-state quantization is governed by several independent constants or so-called quantum numbers which include the following:

n: (principal quantum number); $n = 1, 2, 3.$
l: (azimuthal quantum number); $l \leq (n - 1)$
m_l: (magnetic quantum number); $m_l = -l, -(l-1)\ldots 0\ldots(l-1), l$

where in Fig. 5 m_l is associated with the angle, ϕ, in the x-y plane, l is associated with the azimuthal angle, θ, and of course n is related to r. Consequently, r characterizes the quantum state or orbital distance from the nucleus, where historically for the so-called Bohr atom (after N. Bohr), electrons were considered to be associated with "shells" represented by K, L, M, N, etc. corresponding to the quantum states 1, 2, 3, 4, etc., respectively (Fig. 5b). Note that for spherical symmetry illustrated for the atom geometry in Fig. 4, values of l and m_l for any given value of n have identical energy.

The wave equation representative of a multi-electron atom (above hydrogen as depicted ideally in Fig. 4) can be expressed as

$$\sum_{k=1}^{k=N}\left[\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right)+\frac{2m}{\hbar^2}\left(\frac{e^2Z}{r_k}-V_k\right)\psi+\frac{2m}{\hbar^2}E\psi=0\right], \tag{22}$$

where

$$V_k = \sum_{i=1}^{N}\frac{e^2}{r_{ki}}, \tag{23}$$

and r_{ki} represents the distance between the *k*th and *i*th electron. Like the many-body problem in classical mechanics, Eq. 22 does not yield an exact solution and can be

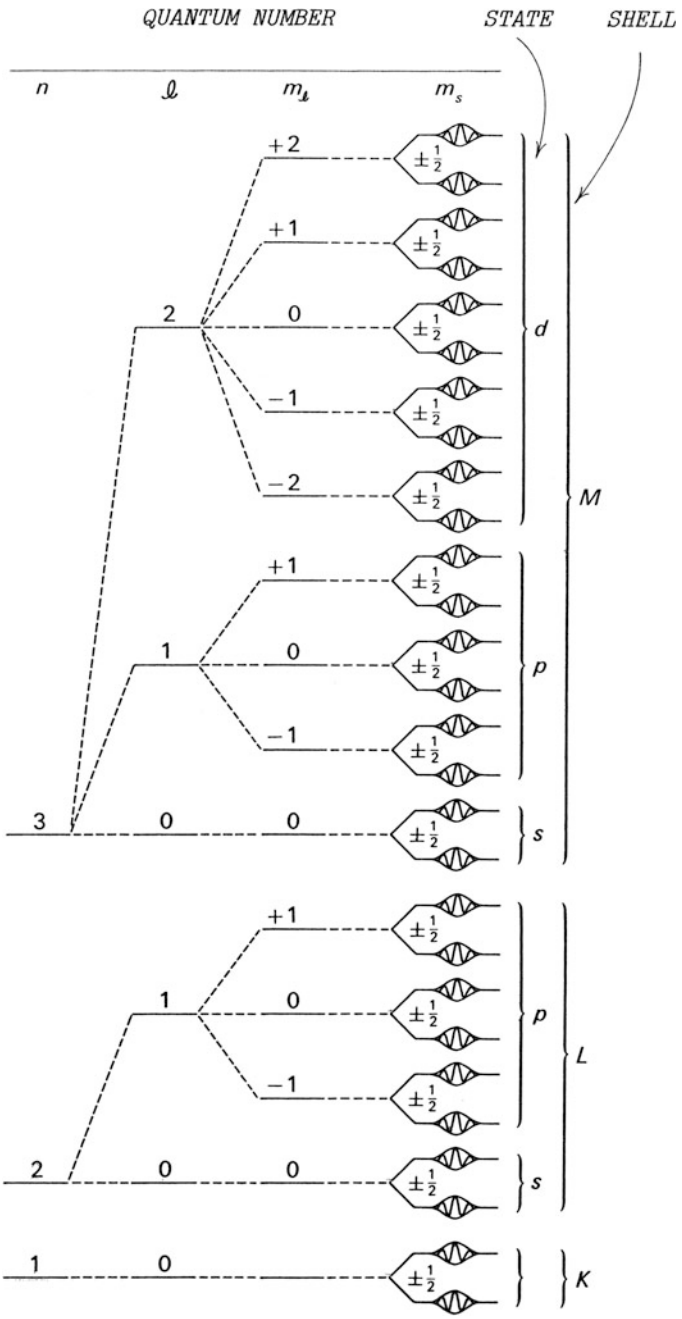


Fig. 6 Atomic energy-level diagram for n = 1, 2, 3. Note the electron pairs imply that each electron spins opposite to the other with reference to the wave packet axis. The quantum numbers are represented in columns: n, l, m_l, m_s

solved in various ways, giving rise to electrons represented in quantized energy levels represented by n = 1, 2, 3, etc. or more specifically for an atom $-1/n^2$ depicted ideally in Fig. 5b.

It might be noted that the atomic quantum levels represented in Fig. 5b have a total number of electrons given by

$$2\sum_{l=0}^{l=n-1}(2l+1) = 2n^2, \tag{24}$$

where the azimuthal quantum number, 1, designates the number of sublevels or substates which as shown vary from zero to n−1. These sublevels are designated s (for l = 0) p (for l = 1), d (for l = 2), and f (for l = 3). Consequently, for n = 1, l = 0 and two electrons (Eq. 23) occupy the s-substate. However, W. Pauli proposed that no two electrons can have the same quantum numbers, and in 1925, G. E. Uhlenbeck and S. Goudsmit postulated that electrons possess a finite spin about some axis. This spin, with a constant angular velocity, has come to be represented by a spin quantum number, $m_s = \pm 1/2$, indicative that the two electrons occupy the 1 s quantum substate (designated 1 s^2), with each electron spinning in opposite directions. Consequently, for each quantum state (n) as shown in Fig. 5b, the s-substate contains two electrons with opposite spin: n = 1 (1 s^2), n = 2 (2 s^2), n = 3 (3 s^2), etc. From Eq. 24, it can be observed that the implicit energy degeneracy implies that for n = 2, the substates are designated as $2s^2 2p^6$, since for n = 2, there are eight electrons in the quantum level (L), requiring six electrons in the p-substate. Consequently, for an atom with the K and L levels filled (n = 1 and n = 2), we can designate this atom as $1s^2 2s^2 2p^6$ representing a total of 10 electrons in two filled quantum levels, n = 1, 2; representing neon (Ne); 2 + 8 electrons. Figure 6 illustrates an energy-level diagram representing the complete quantum structure for atoms implicit in Eq. 24.

References

Moore WJ (1992) Schrödinger: life and thought. Cambridge University Press, Cambridge, UK

Muller-Kirsten HJW (2012) Introduction to quantum mechanics: schrödinger equation and path integral, 2nd edn. World Scientific, Miami

Murr LE (1978) Solid-state electronics. Marcel Dekker, New York

Summarizing Atom and Ion Structure: The Periodic Table of the Elements

Contents

Abstract

Having illustrated symbolically and mathematically the wave-particle dualism applied to electrons in atoms in chapter "A Brief Introduction to Quantum Mechanics", the symmetrical and asymmetrical behavior of quantized electron energy states in composing or characterizing atoms in the development and prediction of atomic chronology implicit in the periodic table of the elements is explored. The systematic or asystematic addition and subtraction of electrons from the quantum states or substates of atoms is discussed as this creates ions, both positive (cations) and negative (anions) valence states. This serves as a precursor to understanding change balance and atomic or ionic binding in the creation of solid or condensed matter to be discussed in chapter "► Chemical Forces: Molecules" following.

Introduction

The concept of atoms as fundamental building blocks for solids (or condensed matter) dates to at least 440 B.C., but the Greeks, in particular Aristotle, were generally opposed to the concept. Plato (circa 330 B.C.) envisioned fundamental categories of "elements" as air, fire, earth, and water, along with the ether or the "universe" to be represented as five polyhedral forms respectively: tetrahedron, octahedron, cube (or hexahedron), icosahedron, and dodecahedron, referred to as the Platonic solids illustrated in Fig. 1. These characterize the initial 5 polyhedral

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_6

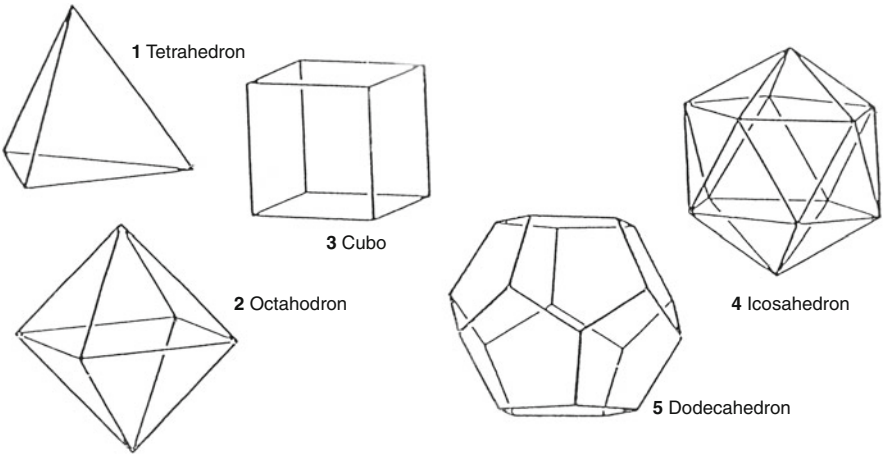


Fig. 1 The five Platonic solids

forms for the regular-faced, convex polyhedra as described in the lifework of Berman (1971). Polyhedra characterize the growth of crystals as nanocrystals or nanoparticles, where the icosahedron and faceting of the cube corners to form a cube octahedron or a truncated octahedron shown in Fig. 2 dominate. The cube octahedron and truncated octahedron shown in Fig. 2a, b represent progressive polyhedral in the sequence of regular-faced convex polyhedral discussed on detail by Berman (1971). The Platonic solids in Fig. 1 represent the first 5 of these while the cube octahedron and truncated octahedron are ranked number 6 and 9, respectively. The regular faces or polygons composing these polyhedral are represented in the associated nets in Fig. 2a, b.

As noted in Part I, metal extraction, pottery making, glass working, pigment development, etc., were all related to chemistry, but early metallurgy as described in De Re Metallica and Alchemy, the protoscience of chemistry, was not able to explain the nature of matter and materials in relation to their processing or fabrication. Robert Boyle in 1661 presented a clear distinction between alchemy and chemistry, whereupon chemistry emerged as a science with the development of the chemical fundamentals such as the conservation of mass by A. Lavoisier (1743–1794), along with thermodynamic principles established later in the work of J. W. Gibbs (1839–1903). Lavoisier and Joseph Priestly, around the same time, almost simultaneously discovered oxygen, and Priestly was also involved in a wide range of both electrical and chemical experiments – a precursor to electrochemistry. His work in fact influenced M. Faraday and Maxwell. Sir Humphry Davy also influenced the work of Faraday in the early 1800s by his publication, "On Some Chemical Agencies of Electricity." Davy's pioneering work in electrolysis allowed him to discover Na, K, Mg, Ba, and B. He also discovered nitrous oxide and its effect as a "laughing gas," but he never considered its use as an anesthetic.

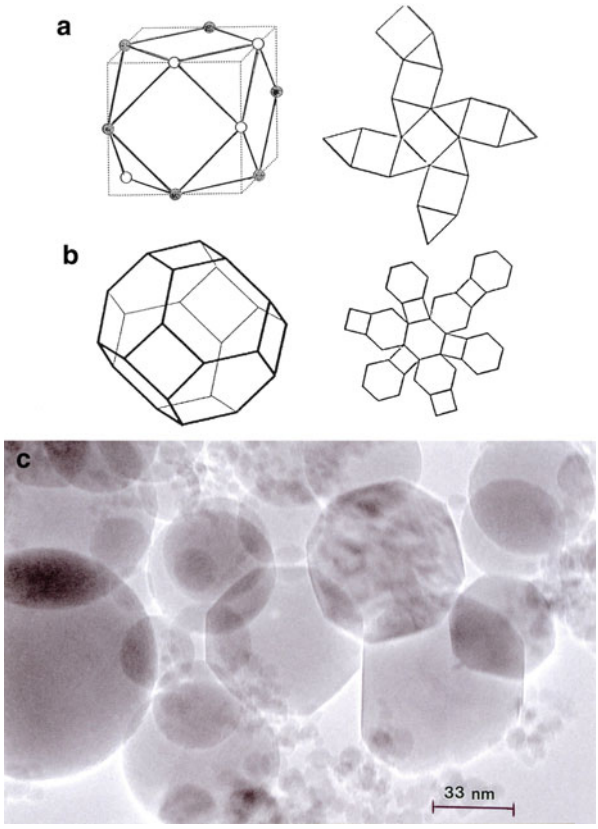


Fig. 2 Nanocrystal growth as cuboctahedra (**a**) or truncated octahedra (**b**). The nets composing the polygons or faces are shown. Note triangles in the cuboctahedron become hexagons in the truncated octahedron. (**c**) Projections of Al_2O_3 particles showing octahedral facets in the transmission electron microscope

In the later 1800s, Dmitri Mendeleev, a Russian chemistry professor, and J. Lothar Meyer, a German chemist, independently published periodic tables of the elements in rows or columns according to their atomic weight, starting new rows or columns when certain characteristics repeated. Mendeleev left gaps in his table where it seemed elements had not been discovered, and he was the first to use trends in his chart to predict properties of the missing elements. Gallium and germanium were among the notable elements to be predicted. Mendeleev also occasionally ignored the order suggested by the element's atomic weight. It became apparent that Mendeleev had organized the elements in order of increasing atomic number (Z). Because of the importance of the periodic arrangement of the elements and the ability to predict new elements and their positions in the periodic chart, it might be assumed that Mendeleev would have received a Nobel Prize. While he was in fact nominated in 1906 and 1907 for the Nobel Prize in chemistry, he was rejected on personal bases, in particular his criticism of the work of S. Arrhenius, who pressed for his rejection by the Nobel Committee. In addition, this ordering was consistent with the number of protons which equaled the number of electrons (Z).

Antonius Van der Broek, almost simultaneous with an idea by Ernest Rutherford in 1911, suggested that the central charge and number of electrons in an atom was exactly equal to its place in the periodic table. Consequently, by the time Schrödinger developed his quantum wave equation (▶ Eq. 13 of chapter "A Brief Introduction to Quantum Mechanics"), it was well recognized that atoms were characterized by their number of electrons or atomic number, Z, which balanced the number of protons in the nucleus. Using this hypothesis, Henry Moseley, in 1913, measured the wavelength of X-ray photos for electron transitions produced by elements ranging from Al (Z = 13) to Au (Z = 79) using X-ray excitation and concluded that the atomic number was related to the electron quantum level energy transition, or

$$\Delta E' \propto \frac{1}{\lambda'} = C\,(Z - \sigma)^2, \tag{1}$$

where c is a constant and σ is a nuclear screening factor which increases from ~3 to ~5 for elements with Z = 30 (copper, Cu) to Z = 90 (thorium, Th). This accounts for the increasing shielding or screening provided as the number of electrons and the corresponding electron cloud surrounding the nucleus increases. Equation 1 has become known as Moseley's law and forms the basis for elemental analysis in matter using X-ray energy dispersive spectrometry (EDS) which measures characteristic X-ray energy, E′ (or ΔE′), or wavelength, λ′ according to Eq. 1, quantifying elemental or atomic concentrations in real-time analyses.

Physical Meaning of the Wave Function: Wave Model of Atoms

While we have dealt with the energy of atoms in relation to atomic energy levels and quantum numbers, we have not specifically considered the physical meaning of the wave function, ψ, which represents the wave amplitude. Correspondingly, the notion that an electron beam is considered to be a wave with amplitude, ψ, the beam intensity will be represented by ψ^2 or $\psi^*\psi = |\psi|^2$ where ψ^* is the complex conjugate of ψ. Assuming dV to be an element of intra-atomic volume, we can observe that

$$e\int |\psi|^2 dV = e \quad \text{or} \int |\psi|^2 dV = 1 \tag{2}$$

which expresses the total charge of the *electron cloud* is equal to e. Therefore, $e\,|\psi|^2$ may be considered the charge density of the atom. According to M. Born, the quantity $|\psi|^2\,dV$ is the *probability* of finding an electron in a given volume element, dV. This means that the quantity $|\psi|^2$ becomes the *probability density*, and in this context ψ is the amplitude of the probability. In this case, Eq. 1 indicates that the probability of finding an electron inside an atom is unity, and calculating the quantity $|\psi|^2$ from the wave equation only allows for the probability of finding an

electron at a given point in space. This is expressed by the *Heisenberg uncertainty principle*:

$$\Delta x \cdot \Delta p = \hbar \tag{3}$$

where Δx is the measure of the uncertainty of the geometric position on an electron (in reference to "▶ Fig. 4 of chapter A Brief Introduction to Quantum Mechanics") and Δp is the momentum uncertainty ($p = mv$). The significance of Eq. 3 is that the more precisely we can determine the electron momentum or energy, the less precisely its position can be determined. Correspondingly, since the energy of an electron in the steady state of an atom is fixed as given in "▶ Eq. 20 of chapter A Brief Introduction to Quantum Mechanics." it follows that the "orbit," implicit in ▶ Fig. 4 of chapter "A Brief Introduction to Quantum Mechanics," is impossible to define.

In retrospect, it might be recalled that with reference to "▶ Fig. 4 (chapter A Brief Introduction to Quantum Mechanics") (ideally representing a one-electron atom, the hydrogen atom), the quantum numbers n, l, and m_l are directly related to the positional coordinates describing the orbital position of the electron. The radial position, r, is dependent on n, while l and m_l relate generally to spherical coordinates θ and ϕ. If we imagine a normalized wave function having arbitrarily fixed angles θ and ϕ for a momentarily fixed orbital position for the electron ("▶ Fig. 4 of chapter A Brief Introduction to Quantum Mechanics"), then the r-dependent wave function, $\psi(r)$, for the solution of the Schrödinger equation in spherical coordinates represented by "▶ Eq. 21 of chapter A Brief Introduction to Quantum Mechanics," can be plotted simply as a function of r. The electron orbit can be considered to be established when $|\psi(r)|^2$ is unity for each quantum state (n). Figure 3 illustrates this argument in representing probability distributions or atomic orbital probabilities for helium ($Z = 2$), beryllium ($Z = 4$), and carbon ($Z = 6$) for electrons at a distance r from the nucleus ("▶ Fig. 4 of chapter A Brief Introduction to Quantum Mechanics").

To conceptualize a multi-electron atom requires the superposition of probability distributions shown in Fig. 3 visualized in 3 dimensions. As electrons are added, this complex distribution expands, creating a dense "cloud" which increasingly shields the nucleus (Eq. 1). The important features of the atoms or elements are therefore represented by the outer reaches of this complex, 3-D, probabilistic electron cloud. When viewed, the atom or atoms composing solid matter, especially crystals composed of regular arrays of atoms, will appear as arrays of "cotton balls." This is in fact what is observed in various modes of visualization or atomic imaging shown in Fig. 4 for charged volumes on a crystal surface viewed in the field ion microscope in Fig. 4a in contrast to electron clouds imaged in the transmission electron microscope (TEM) where an electron beam interacts with the electron cloud characterizing each atom in Fig. 4b. The TEM image in Fig. 4a highlights rows of atoms composing specific crystal planes, specifically the {002} planes in fcc gold.

Actually, the notion of a dense electron cloud of probabilistic electron shells surrounding the nucleus implied in the atom images shown in Fig. 4 is somewhat

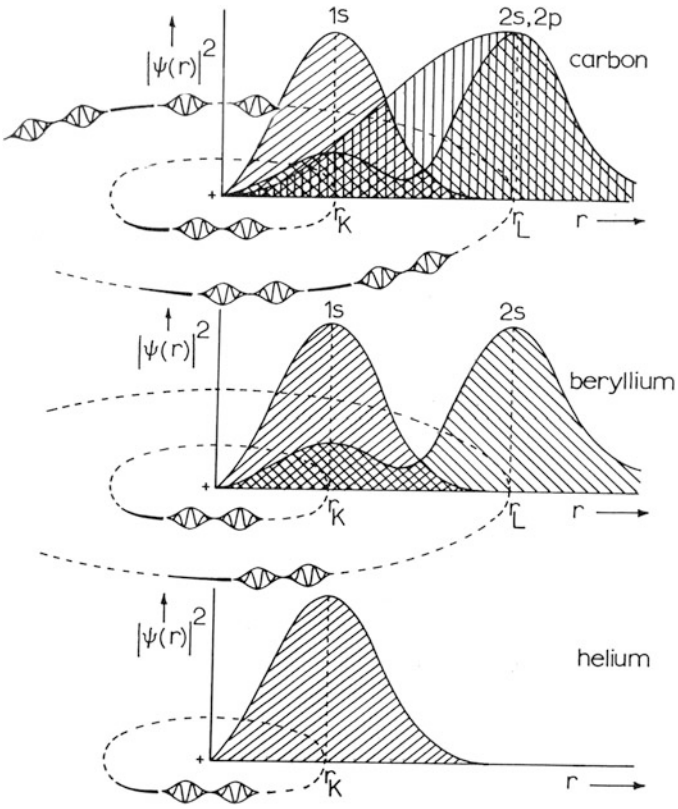


Fig. 3 Atomic orbital probabilities represented only as radial functions, r_K, r_L. (not to any scale)

misleading since from a spatial perspective, the actual size of an electron and the orbital radii, r_K, r_L, etc., in contrast to the corresponding size of the nucleus represents this regime as mostly empty space, just as the milky way galaxy and the universe in total represent mainly empty space relative to the sizes of the component planets, moons, galaxies, stars, etc. Collectively these complex concepts provide a physical model of a quantum-mechanical atom.

The Periodic System of the Elements: Periodic Table

The representation of electrons in atoms as quantized states, along with the Pauli exclusion principle and the fact that each atom is characterized by a least-energy ground state represented by the K-shell or quantum number $n = 1$, provides proof that atoms or elements follow a periodic system represented ideally in "► Fig. 6 of chapter Electromagnetic Fundamentals." As shown in "► Eq. 24 of chapter A Brief Introduction to Quantum Mechanics," the maximum number of electrons in the

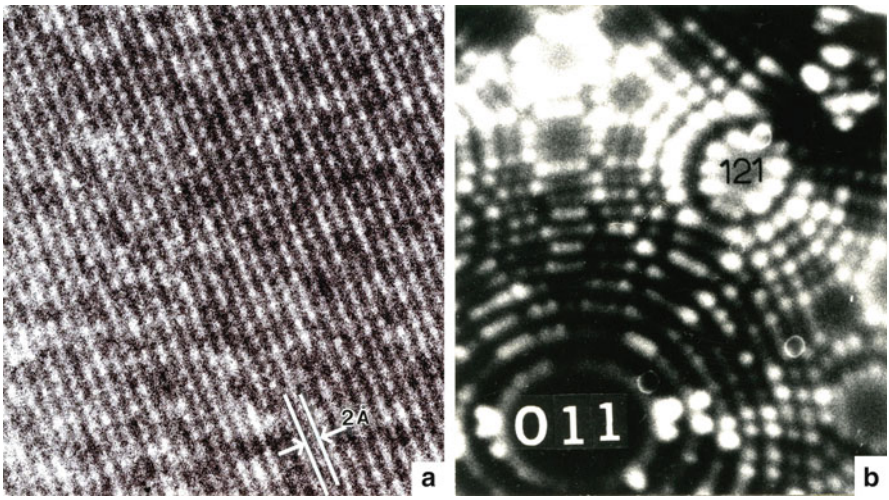


Fig. 4 Images of atoms. (**a**) High-resolution image of gold atoms or atom planes viewed edge-on in a thin gold film in the transmission electron microscope. (**b**) Atom sites observed as helium ions on electric field lines from surface atoms at the edges of atomic planes of tungsten (Z = 74) in a field ion microscope (FIM)

n-quantum number electron groups is $2n^2$. These electrons are distributed among n-subgroups with maximum number of electrons in each subgroup given by 2 (2 l + 1). This number is determined by the possible values of the quantum number m_l (for given values of l) and by the $\pm$ 1/2 for M_s. This diversity of electron states and substates (as represented graphically in "▶ Fig. 6 in chapter A Brief Introduction to Quantum Mechanics") can be written as:

K(n = 1)	**L(n = 2)**	**M(n = 3)**	**N(n = 4)**	**O(n = 5)**
1s	2s	3s	4s	5s
	2p	3p	4p	5p
		3d	4d	5d
			4f	5f
				5g

In state n = 1 (1s) there are 2s electrons (since l = 0) forming the two-electron K-shell or group of the atom ($1s^2$). In state n = 2, we have 2s electrons and 2 (2 × 1 + 1) = 6 p electrons ($2p^6$) since l = 1 in "▶ Eq. 24 of A Brief Introduction to Quantum Mechanics." This produces 8 electrons in the L (n = 2) group, consisting of the subgroups 2s and 2p as shown above. For the M-shell (n = 3), there are three subgroups which fill as follows: $3s^2 3p^6 3d^{10}$ since the total number of electrons in the M-group is 18. However, experiment shows that the maximum number of electrons for the M-group (18) is only found in heavier elements beginning with copper, Z = 29. In elements following Ca (Z = 20), the 4s and 3d quantum substates overlap, and this produces a quantum state filling

anomaly beginning with scandium (Sc) (Z = 21) and ending with Cu (Z = 29). Elements in this anomalous series are referred to as transition elements and possess unique properties such as magnetism (especially Fe (Z = 26). After copper, the systematic electron filling of subgroups continues up to Z = 36 (Kr). For krypton (Kr) and the next two elements (Rb: Z = 37 and Sr: Z = 38), the N-group stops at 8 electrons because their outer electrons occupy the 5s substate and not the 4s substate which continues another anomalous series up to palladium (Pd: Z = 46) where the number of N-state electrons reaches 18. Other anomalous groups include elements identified as rare earths (cerium, Ce: Z = 58 to lutetium, Lu: Z = 71) and the actinides: thorium, Th (Z = 90) to californium, Cf (Z = 98). These features are represented in the periodic table of Fig. 5 which illustrates valence states, or the charge of the atom depending upon the loss or gain of electrons (+or, respectively). Atoms (or elements) with the same valence number or charge are generally arranged in columns beginning with positive ions (or cations) with charge +1, beginning with hydrogen, followed by the next column of positive ions with charge +2, beginning with beryllium at the left of the table. The extreme right column contains the rare gas elements which ideally have zero valence, beginning with helium (Z = 2).

An examination of the electronic structure represented in the periodic table of Fig. 5 can provide a predictive matrix for compound formation (chemistry) especially for ionic or electrovalent solids or ionic crystals represented ideally in "▶ Fig. 10 of chapter Electromagnetic Color and Color in Materials" for NaCl (Na^+Cl^-) and ▶ Fig. 21 of chapter "Examples of Materials Science and Engineering in Antiquity" for CaF_2 ($Ca^{2}+ F_2^-$). Note in Fig. 5 that elements to the left of the rare earth column usually represent anions (having a negative charge) since these elements forming ionic molecules accept electrons. Those columns headed by F (Z = 9) and O (Z = 8) have a valence of 1- and 2-, respectively, although other possible valences are noted. Ionic or electrovalent compounds are therefore formed by the transfer of electrons from elements in columns to the left to element in columns to the right in the periodic table. In the case of covalent (shared electron) bonds, the theory of electron pairs represents the most important generalization of the results of quantum-mechanical analysis of simple molecules such as C Cl_4 which can also be represented electrovalently as $C^4 + Cl_4^-$. Similarly as a covalent compound, CH_4 can also be represented electrovalently as $C^{4-} H_4^+$ or CO ($C^{2+}O^{2-}$) as illustrated in Fig. 5 where carbon is represented by a valence of 2 or 4. However, for covalent binding which prevails, the 4 electrons in 2p for carbon are shared with 3p in Cl for C Cl_4, stabilizing 2s ($2s^2$) for carbon and 3p ($3p^6$) for chlorine. In this case, we can examine the carbon outer electron configuration: $2s^22p^2$. Carbon can also stabilize its electronic structure in the $2s^2$ substate by giving 2 electrons in the $2p^2$ substate filling. This corresponds to a valence or charge of +2. Correspondingly, carbon can receive 4 electrons in the p-substate to stabilize (or fill) $2p^6$, having an electrovalence of −4. In contrast, carbon can give up the $2s^22p^2$ electrons to stabilize the ion in the $1s^2$ substate, corresponding to a valence of +4. It is this quantum state or substate filling which forms stable and even energetically favorable pairing of elements forming molecules. This is

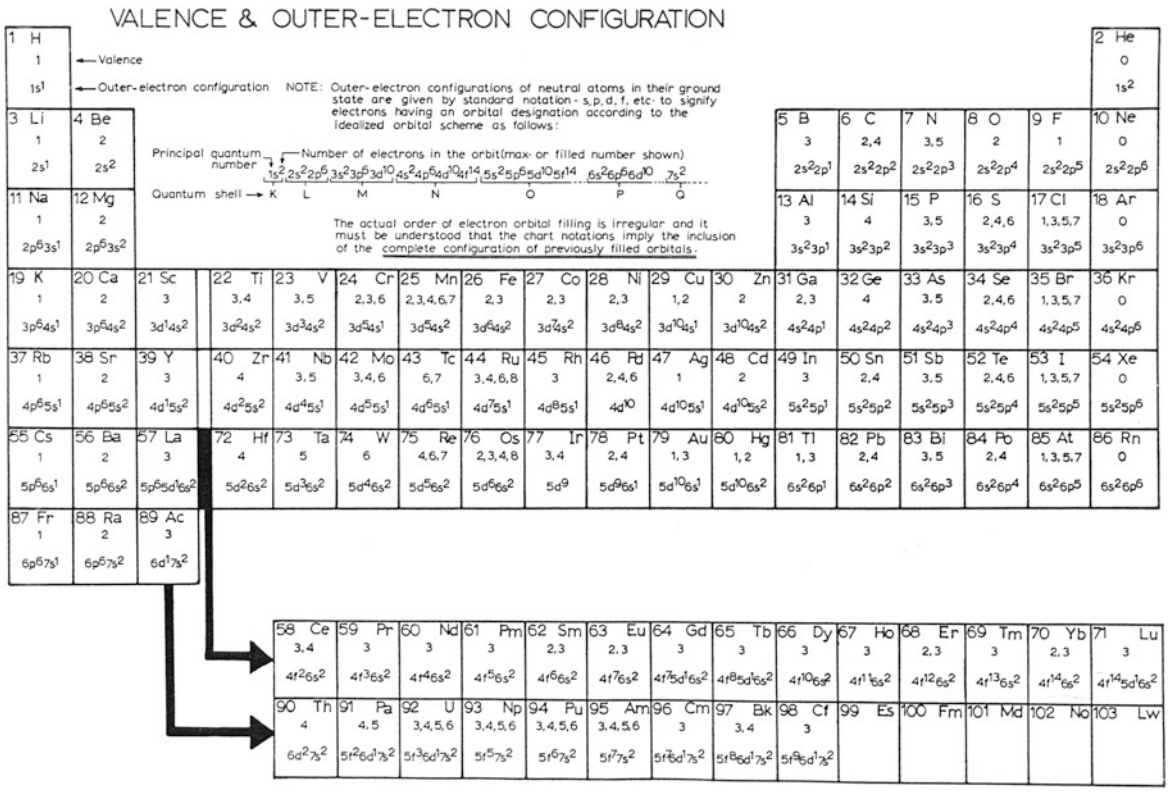


Fig. 5 Periodic table of the elements showing valence and electron configurations (From Staudhammer and Murr (1973))

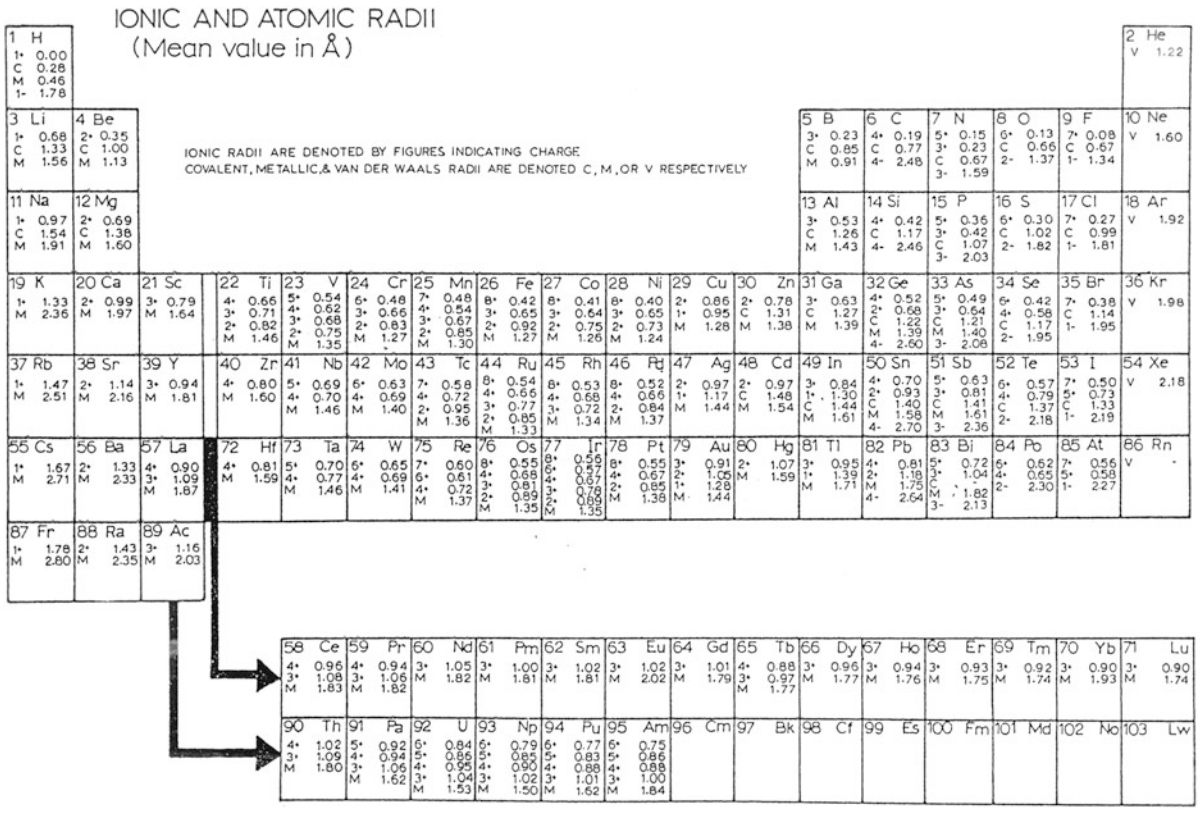


Fig. 6 Periodic table of the elements showing ionic and atomic radii (From Staudhammer and Murr (1973))

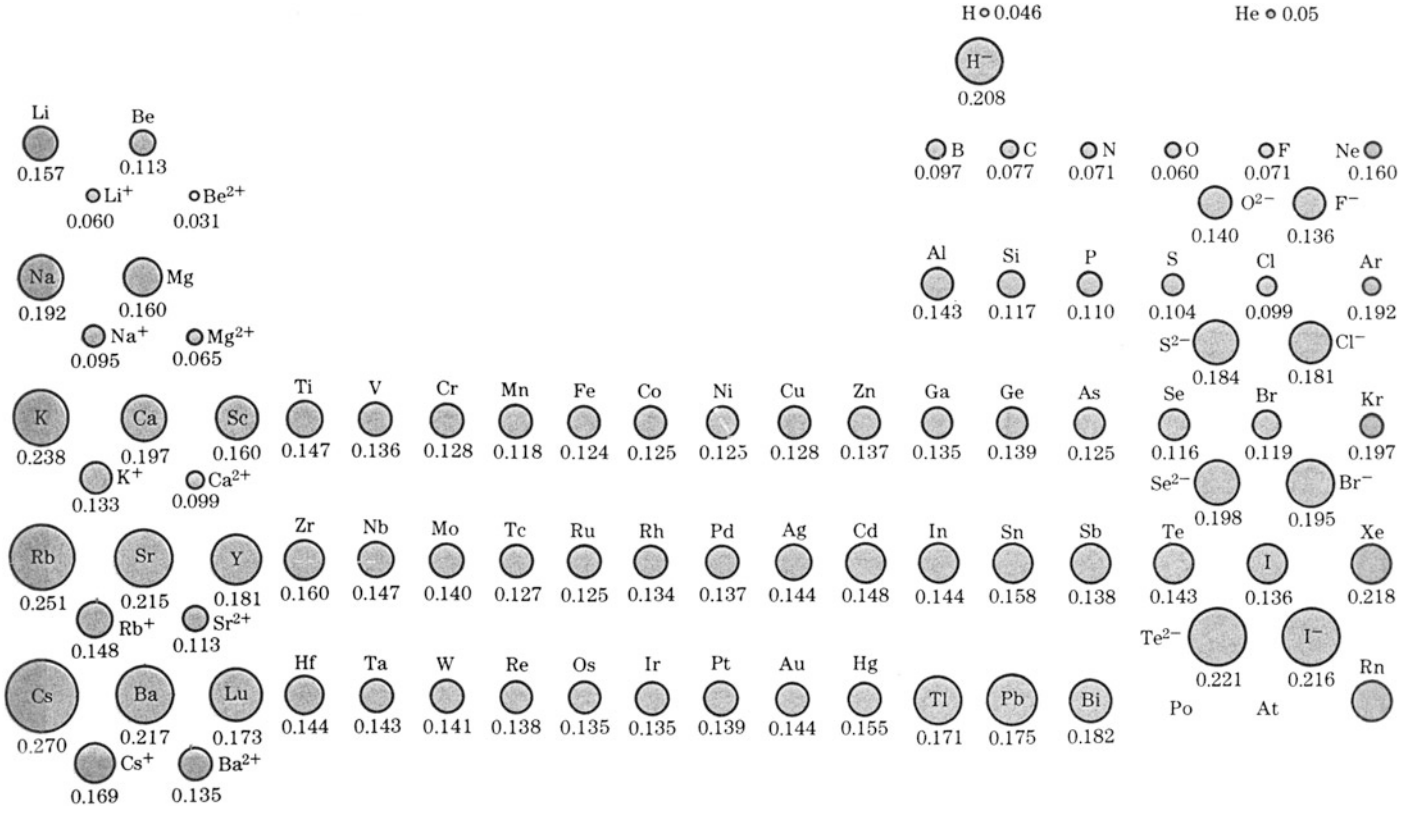


Fig. 7 Periodic table of the elements showing relative sizes for some atoms (neutral atoms) and ions. Sizes noted one radii in nanometers. Ions are shown by charge (+ or −) (Adapted from Smith (1990))

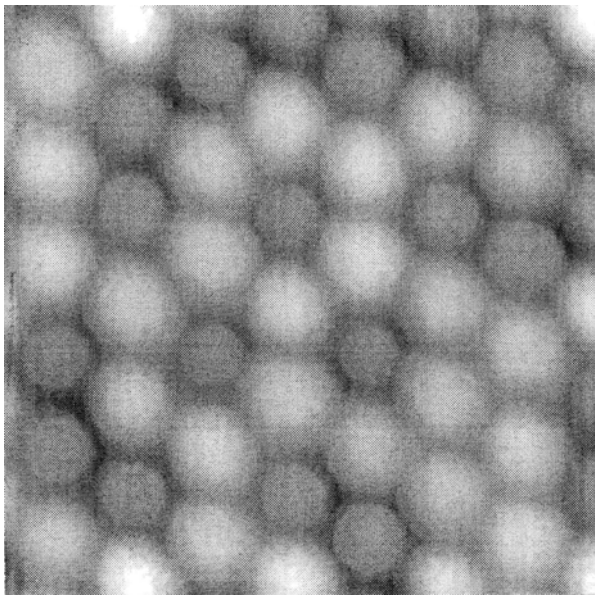


Fig. 8 Atomic force microscope image differentiating Si and Fe atoms in a mixed compound (Courtesy of Oscar Custance, University College, London)

demonstrated by the diamond structure in "▶ Fig. 7 of chapter Electromagnetic Color and Color in Materials" where the strong covalent bond is formed on sharing 4 electrons with 4 carbon atoms, but two carbon atoms can be ionically stabilized by $2s^2 2p^2$ going to stabilize the 2p substate with 4 electrons. In this case the valence becomes 4+ and 4−. $CaCO_3$ ($Ca^{2+}C^{4+}O^{2-}$) is a more specific ionic or electrovalent compound example.

The most stable electronic configurations are observed to be the filling of the nsnp quantum substates: n = 1 (2 electrons in $1s^2$), n = 2 (10 electrons in $1s^2 2s^2 2p^6$), n = 3 (18 electrons in $1s^2 2s^2 2p^6 3s^2 3p^6$), etc., which is represented by the rare (or inert) gases in the extreme right column in the periodic table of Fig. 5. Considering the nsnp filling or stabilization rule can also determine the normal or preferred valence states. For example, for the column to the left of the inert gases, the valence is 1-, while for the next column it is 2-. But, as noted, the column represented by C and Si can have mixed valence since fewer electrons are involved.

The gain or loss of electrons by an atom creating positive (+) or negative (−) valence (or electrovalence) must correspondingly cause an increase or decrease in the resulting ion size with respect to the neutral atom having 2 electrons. This feature is illustrated in the corresponding periodic table shown in Fig. 6. The atom sizes for neutral atoms usually imply metal (metallic) elements which are characterized by metallic binding (M in Fig. 6). Covalently bound atom sizes are also noted in Fig. 6 by C as noted. Weak binding atom sizes are also denoted V representing van der Waals binding. We will discuss binding in more detail in Chapter 7 following. Figure 6 presents a scaled representation for atom or ion (+or −) sizes (radii) corresponding to Fig. 7. It can be observed in this relative size

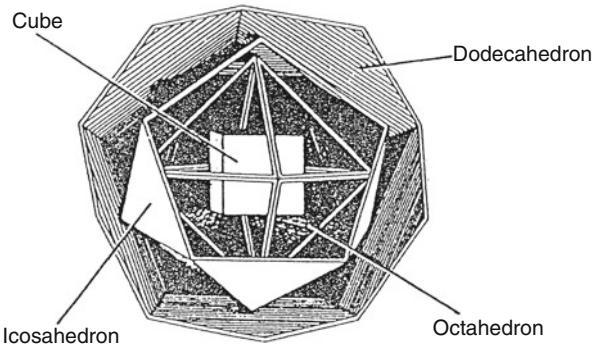


Fig. 9 Sequence of four Platonic polyhedral whose 46 vertices form half of the "Keplerian atom"

representation of the periodic table of the elements that the columns of elements at the left and right of the table show increasing size with increasing Z in each column (going down the columns) while the sizes also generally increase in elemental rows from left to right, also representing increasing Z from left to right in each row.

Figure 8 illustrates size and density variations detected in an Si–Fe compound by atomic force microscopy. Compare Si and Fe radii as represented in Fig. 6, for example.

It is of interest to point out that the representation of "elements" invested in the Platonic solids (more than 2300 years ago) shown in Fig. 1 have been configured in more recent times to represent elements from $Z = 1$ to $Z = 92$ by nesting, or systematically and geometrically placing one or more of the Platonic polyhedra within one another. Simple Platonic duals, where one polyhedron can be placed within the other and vice versa, include the cube and octahedron and the tetrahedron and the icosahedron and dodecahedron. Four nested Platonic polyhedra including the icosahedron inside the dodecahedron, the octahedron inside the icosahedron, and the cube (or hexahedron) inside the octahedron contain 46 vertices which form half of what is referred to as the "Keplerian atom" and illustrated in Fig. 9. This concept was originated by Robert Moon in the mid-twentieth century. It forms a geometric basis for the periodicity of the elements to create an electron shell model using Platonic solids and their duals, commonly referred to as the Moon Model (Hecht 2000; Hecht and Stevens 2004). In some respects, this polyhedral nesting also emulates shell and cluster formation to be discussed in chapter "▶ Chemical Forces: Nanoparticles".

References

Berman M (1971) Regular-faced convex polyhedra. J Frankl Inst 291(5):329–352
Hecht L (2000) Advances in developing the moon nuclear model. 21st Century Fall:5–12
Hecht L, Stevens CB (2004) The moon model. 21st Century Fall:58–74
Smith WF (1990) Principles of materials science and engineering. McGraw-Hill, New York
Staudhammer KP, Murr LE (1973) Atlas of binary alloys: a periodic index. Marcel Dekker, NewYork

Chemical Forces: Molecules

Contents

Abstract

The way atoms and ions are bound creates many of the properties of solid matter, including electrical, mechanical, and optical properties in particular. These chemical forces include ionic, covalent, metallic, and van der Waals binding, which are briefly but fundamentally described in this chapter.

Introduction

Alexander Butlerov, a Russian chemist, in 1861 created the theory of chemical structure, the principle of which is as follows: "starting with a concept that each *chemical* atom that makes up a body participates in the formation of this body and operates via specific quantity of force inherent in it (affinity). I call *chemical structure* a distribution of action of this force, as a result of which the chemical atoms (acting upon one another directly or indirectly) combine into a chemical particle . . . the chemical nature of a complex particle, is determined by the nature of the elementary component parts, their number, and chemical structure."

While we can associate the chemical elements as defined in the periodic arrangements shown as depicted in the periodic tables of ► Figs. 5, ► 6, and ► 7

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_7

in chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements" with crystal structures and corresponding bonding forces, covalent crystals (such as H, B, C, Si, Ge) and metallic crystals (such as Li, Be, Na, Mg, Al . . . Ca to Zn, etc.), molecules, and even single molecules such as NaCl (Na^+Cl^-) representing ionic molecules and H_2 (H:H) representing covalent molecules, along with weakly bound molecular arrays (van der Waals molecules), constitute other crystal structures. For example, typical ionic crystals such as NaCl-type structures and CaF_2-type structures as shown in ▶ Figs. 10 and ▶ 11 of chapter "Electromagnetic Color and Color in Materials" involve elemental mixtures. Van der Waals forces are apparent in graphite, the hexagonal crystal structure for carbon where the hexagonal carbon lattice, represented by the graphene lattice in ▶ Fig. 5 (chapter "Examples of Materials Science and Engineering in Antiquity"), is weakly bound in the c-axis direction of the graphite crystal structure while in the a-b plane of the graphene lattice (▶ Fig. 5 of chapter "Examples of Materials Science and Engineering in Antiquity"), the bonding is covalent, a strong force. Metallic bonding exists for all of the regular crystal structures which can involve a complex mixing of natural valence metallic elements producing complex chemistries for metal alloys, while complex metal alloys can also exist in aperiodic crystal structures as quasicrystals. As noted earlier, electrovalent (or ionic) compounds involve valence electrons which are transferred from one element to the other, while for covalent compounds, electrons are shared to produce strong bonds. This bonding forms the basis for chemical bonding or quantum chemistry whose basis involves the Schrodinger wave equation applied to molecules.

The Ionic Bond

In the ionic bond as it applies to charged elements (cations (+) and anions (−) such as Na^+ and Cl^- forming a molecule of Na^+Cl^-, the potential energy can be expressed in the form

$$u = -\frac{Ae^2}{r} + \frac{Be^2}{r^n}, \tag{1}$$

where the first (negative) term is the attractive force potential that expresses the mutual potential energy of the oppositely charged cation and anion as point charges, while the second (positive) term represents the repulsive force potential. In the NaCl crystal, the equilibrium radius or distance between Na^+ and Cl^- can be determined from the lattice parameter or unit cell dimension (5.65 Å) implicit in the NaCl crystal structure in ▶ Fig. 10 of chapter "Electromagnetic Color and Color in Materials" or ideally by setting

$$\left(\frac{du}{dr}\right)_{r=r_o} = 0 \tag{2}$$

where r_o is the crystal lattice constant or a related parameter in the unit cell of ▶ Fig. 10 in chapter "Electromagnetic Color and Color in Materials." From this condition we have

$$B = \frac{A}{n} r_o^{n-1} \tag{3}$$

from which the crystal lattice energy (Eq. 1) becomes

$$u = -\frac{Ae^2}{r}\left[1 - \frac{1}{n}\left(\frac{r_o}{r}\right)^{n-1}\right], \tag{4}$$

and the equilibrium energy becomes equal to

$$u_o = \frac{n-1}{n}\left(\frac{Ae^2}{r_o}\right), \tag{5}$$

where in the NaCl-type crystal structure shown in ▶ Fig. 10 in chapter "Electromagnetic Color and Color in Materials," $r_o = a/2$, where a is the lattice parameter (5.65 Å) and $r_o = 2.83$ Å. A in Eq. 5 is a constant called the *Madelung constant,* which for NaCl-type crystals is of the order of unity, and deviates from unity for other crystal structures such as the CaF_2 type shown in ▶ Fig. 11 in chapter "Electromagnetic Color and Color in Materials."

The Covalent Bond

The first step in applying quantum mechanics to chemical bonding was the configuration of the wave equation (▶ Eq. 19, of chapter "A Brief Introduction to Quantum Mechanics") to approximate the hydrogen molecule (H_2):

$$\sum_{i=1}^{i=2}\left(\frac{\partial^2\psi}{\partial {x_i}^2} + \frac{\partial^2\psi}{\partial {y_i}^2} + \frac{\partial^2\psi}{\partial {z_i}^2}\right) + \frac{2m}{\hbar^2}(E - V)\psi = 0, \tag{6}$$

where

$$V = e^2\left(\frac{1}{r_{12}} + \frac{1}{r}\right) - e^2\sum_{i=1}^{i=2}\left(\frac{1}{r_{ai}} + \frac{1}{r_{bi}}\right) \tag{7}$$

is the potential energy of the molecule (H:H).

In Eq. 7, r_{12} is the distance between the two electrons, r is the distance between the two nuclei (or protons), and r_{ai} and r_{bi} denote the distances of the ith electron (1 or 2) from the (a) nucleus and the (b) nucleus. The total energy for the hydrogen molecule, E, in Eq. 6 (two-atom molecule) can be solved approximately as the

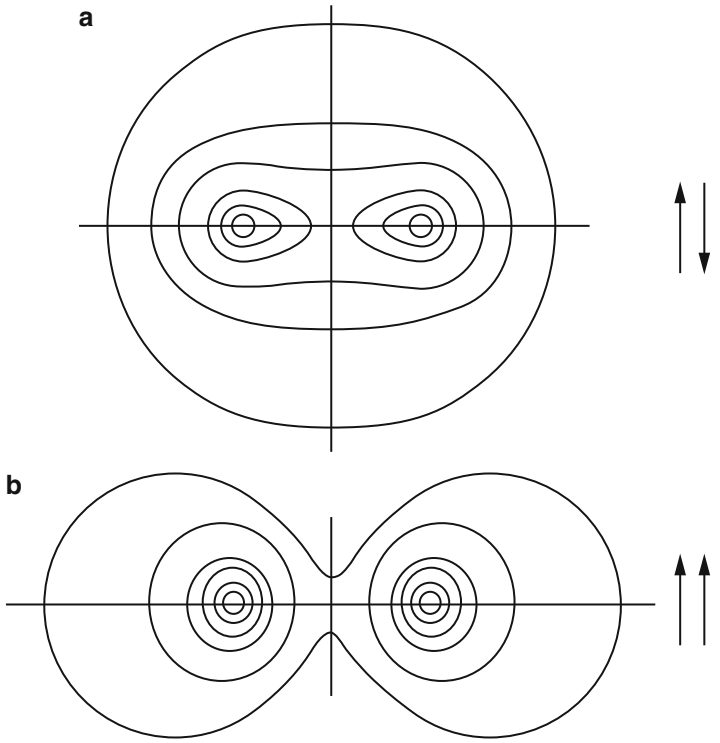


Fig. 1 Distribution of charge density of electron cloud in the stable (**a**) and unstable. (**b**) State of a hydrogen molecule

two-electron atom (helium). Calculating the probability density $|\psi|^2$ for the electron cloud in the hydrogen molecule, it may be shown that for the stable state of the molecule, like the stable helium atom, the electron cloud is concentrated between the nuclei, whereas in a corresponding unstable state, the electron cloud tends to move away from the center of the molecule. Simply stated, the stable state involves electrons with opposite spin as in helium, while for electrons with the same spin, the molecule is unstable. This is illustrated schematically in Fig. 1, which represents a pair of electrons with antiparallel spins denoting attraction while a pair with parallel spins denotes repulsion. For the stable configuration in Fig. 1a, the equilibrium distance between protons has been measured to be 0.74 Å. Hydrogen has been solidified below 3 K and in solid form assumes a hexagonal crystal structure, while at about 1.5 K it becomes diamond cubic (a = 5.29 Å) especially in thin-film form (Curzon and Mascall 1965). This atomic hydrogen structure contains eight atoms.

The stability of electron pairs with antiparallel spin as noted for the hydrogen molecule in Fig. 1a represents the *method of electron pairs* which can be extended to other more complex molecules where valence schemes in each pair of atoms with antiparallel electron spins is connected by a line as is common practice in chemistry

to denote single chemical bonding. For example, while the electronic characters of the inert gas atoms are built up of pairs of electrons with antiparallel spins, like He (the pair basis), this pairing feature is also implicit in the electron configuration shown in Fig. 1. Correspondingly, stable configurations are often observed for antiparallel spins, represented by H–H, Cl–Cl, etc.; however, even in the case of the H_2 molecule, in addition to the covalent scheme of bonding represented by H–H, we might also consider H^+-H^- or H^--H^+, where the wave function would be represented by

$$\psi = a\psi_{at} + b\psi_{ion}, \tag{8}$$

where for an antiparallel electron pair, $\psi_{at} = \psi_a(1)\ \psi_b(2) \pm \psi_a(2)\ \psi_b(1)$ which corresponds to the binding of the second electron to nucleus a and the first to nucleus b, while ψ_{ion} is a wave function corresponding to ionic situations. For an approximate analysis of the H_2 molecule, the energy due to ionic structure of the molecule is small. However, for other molecules or molecular or atomic structures, the contribution of ionic valence, or iconicity, to the energy of the chemical bond may become appreciable. What this means is that more molecules or crystals may be considered purely ionic or electrovalent, while others may be purely covalent, with a range of mixed ionic-covalent bonding or some degree of ionicity. Moreover, for neutral atoms forming metal crystal structures, these valence features are often absent, although some alloy phases can exhibit metallic behavior or metallic properties, such as conductivity characteristic of metallic bonding while simultaneously exhibiting other properties characteristic of covalent binding. As illustrated in the periodic tables in ▶ Figs. 5 and ▶ 6 of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements," there is a clear relationship between electropositive and electronegative valence and the positions of atoms or elements. As we will see later, elements in the same groups or columns tend to have the same crystal structure and some apparent properties at room temperature. Electropositive elements are shown in columns to the left and electronegative elements to the right. Binding forces become more covalent between certain neighboring atoms at the right of the periodic table (▶ Figs. 5 and ▶ 6 of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements") where a true covalent bond is associated with diatomic molecules of the halogens: F_2, Cl_2, Br_2, and I_2.

Hybridization

Carbon is one of the most versatile elements in the periodic table since it forms virtually an infinite number of compounds having single, double, or triple bonds. This bonding versatility is referred to as *hybridization* and is illustrated in Fig. 2. The representation of the p substate shows orbitals for paired electrons in ▶ Fig. 6 of chapter "A Brief Introduction to Quantum Mechanics" where $m_l = -1, 0, +1 => p_x, p_y, p_z$, respectively. For carbon, this ground state configuration can be expressed: $1s^2\ 2s^2 2p_x^1 2p_y^1$. Figure 2a shows the normal valence structure for a

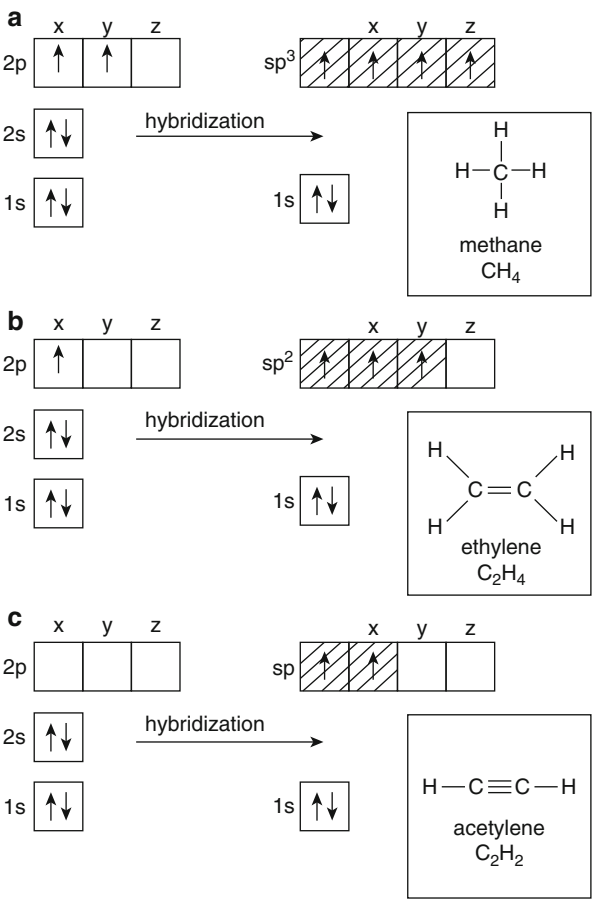


Fig. 2 Schematic showing carbon hybridization: (**a**) sp3, (**b**) sp2, (**c**) sp. Bond angle is 109.5° in (**a**), 120° in (**b**), and 180° in (**c**)

carbon atom where one electron with spin up (↑) occupies p_x and p_y. This electron configuration can be rearranged by putting 1 2s electron into the empty $2p_z$ subshell which results in the 2s and 2p subshell all half filled, resulting in a relatively stable excited state. By sharing an electron in each of these substates, a very stable covalent structure results. This is shown, for example, for methane, CH_4, and is referred to as sp^3 hybridization. The actual bond angle between the carbon and the hydrogen atoms is 109.5, forming a tetrahedron.

In Fig. 2b, p_z is left empty and the hybridization is referred to as sp^2 and illustrated by ethylene, C_2H_4, with a double bond between carbons. Compounds containing double bonds are therefore sp^2 hybridized. Similarly, as shown in Fig. 2c for acetylene, C_2H_2, a triple bond forms between carbon atoms for sp hybridization. In these three hybrid valence states, there are zero, 1, and 2 unfilled p substates corresponding to sp^3, sp^2, and sp hybridization, respectively.

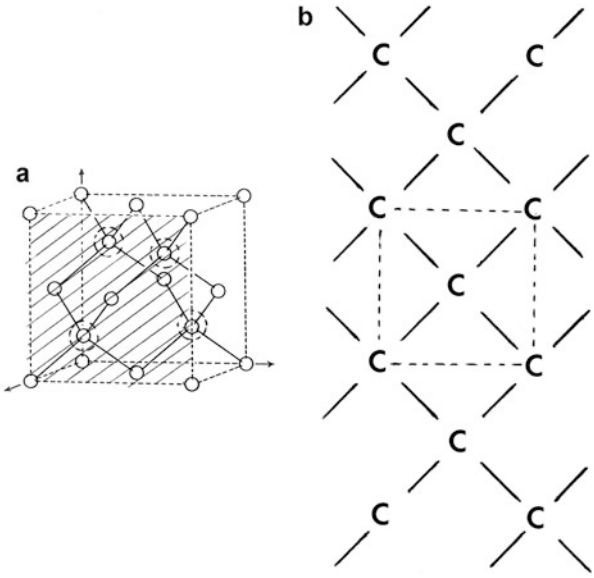


Fig. 3 Diamond cubic unit cell (a) and face-plane schematic showing sp3 hybridization for carbon (b)

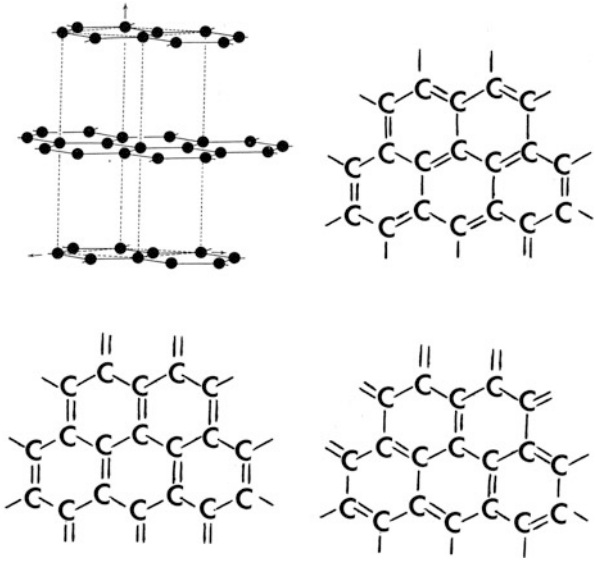


Fig. 4 Hexagonal (unit cell: *dotted lines*) structure and sp2 hybridization variations for carbon (graphene) sheets

Carbon structures, particularly diamond and graphite, are examples of sp^3 and sp^2 hybridization as shown in Figs. 3 and 4, respectively. In Fig. 3a the tetrahedral bonds are more obvious for the clinographic projection showing the interior carbon atoms which are hatched in contrast to the face-centered and corner carbon atoms

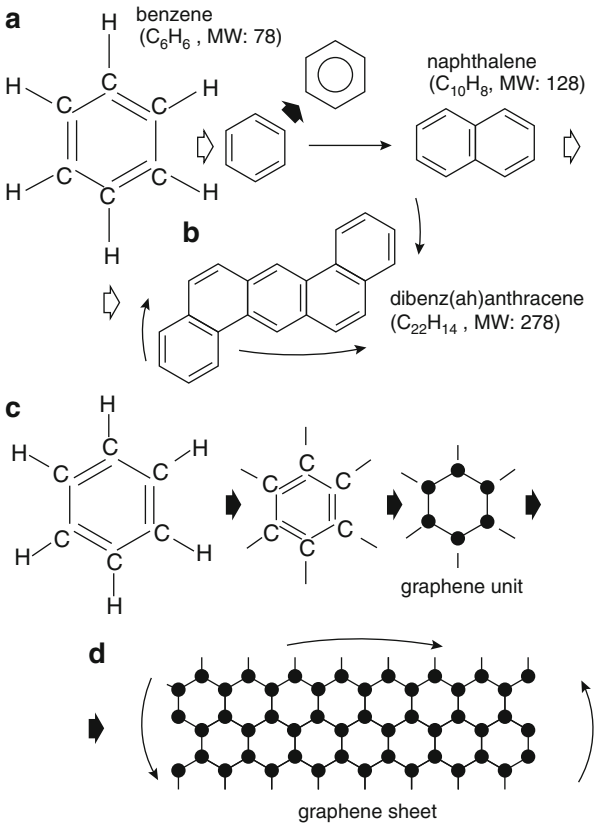


Fig. 5 Benzene sp3 hybridization and derivative structures (Adapted from Murr (2012))

which are open. Shown as a face-plane structure, the sp^3 hybridization is consistent with Fig. 2a within the dotted unit, which also represents the four corner atoms and the face-centered atom in the diamond cubic unit cell projection in Fig. 3a.

In Fig. 4, the clinographic projection of the hexagonal graphite structure shows the hexagonal, graphene layers composing this structure which are weakly bonded by a van der Waals bonding. This sp^2 hybridization is apparent from the three equivalent configurations shown. These equivalent graphene structures can be shown to be derivatives of benzene (C_6H_6) as illustrated in Fig. 5. In Fig. 5a, benzene equivalent structures are illustrated by large arrows while examples of benzenoid compounds are shown as ring combinations and planar ring geometries for naphthalene ($C_{10}H_8$) in Fig. 5a, and dibenz(ah)anthracene ($C_{22}H_{14}$) in ▶ Fig. 19b of chapter "Electromagnetic Fundamentals." These benzene ring structures are referred to as polycyclic aromatic hydrocarbons (PAHs). These structures (isomers) are sp^2 hybridized. In Fig. 5c the removal of hydrogen forms the hexagonal graphene or graphite layer unit which is translated into a graphene sheet in Fig. 5d, corresponding to the 2D graphene lattice discussed in ▶ Fig. 5 of chapter "Examples of Materials Science and Engineering in Antiquity."

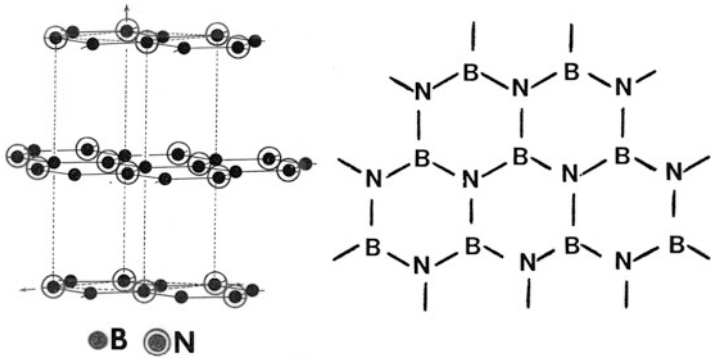


Fig. 6 Hexagonal (unit cell: *dotted lines*) structures and sp^3 hybridization for BN (graphene-like) sheets

Like carbon, silicon ($1s^2 2s^2 2p_x^2 p_y^2 p_z^2 3s^2 3p_x^1 p_y^1$) in the diamond cubic crystal structure is sp^3 hybridized (► Fig. 8, of chapter "Electromagnetic Color and Color in Materials"); SiO_2 and SiH_4 are also similarly sp^3 hybridized along with SiC. BN ($1s^2 2s^2 2p_x^1$: $1s^2 2s^2 2p_x^1 p_y^1 p_z^1$) is diamond cubic (*zinc blende* (ZnS) structure) with sp^3 hybridization but can form a graphite-like hexagonal structure with graphite-like (graphene) layers and corresponding sp^2 hybridization as shown in Fig. 6. In graphite (Fig. 4) the carbon atoms, after sp^2 hybridization, have a p_z orbital containing one electron, while in BN the corresponding p_z orbital is vacant in the boron atom while occupied by two electrons in nitrogen. This difference results in electrical conductivity in graphite, while BN is nonconducting.

Metallic Binding

In contrast to electrovalent and covalent bonding, the bonding in metals is very different. While electrovalent and covalent bonding is valence band related, metallic binding is not. This contrast is provided in the schematics in ► Figs. 6 and ► 7 in chapter "Electromagnetic Color and Color in Materials" where for metals the valence and conduction bands overlap, while for small energy gaps between these bands semiconduction occurs. Conduction diminishes and becomes inconsequential for large energy gaps. Simple metals like aluminum can be considered to have sp binding or sp electrons (compare the Al electronic structure, $1s^2\ 2s^2 2p^6 3s^2 3p^1$, with the sp hybridization in carbon shown in ► Fig. 16c of chapter "Electromagnetic Fundamentals"). Electrons from unfilled sp are detached from the metal ion and become "free." This concept of free electrons embodies the initial work of Drude and Lorentz who proposed an assemblage of positive ions immersed in a matrix (envisioned as a "gas" or "fluid" regime) of electrons not bound to the ion cores. A potential gradient exists at the "surface" of the metal which constrains the electrons

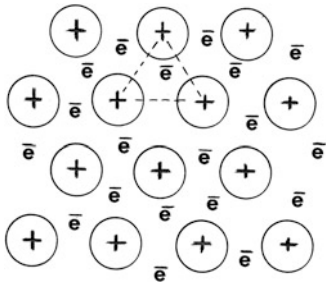


Fig. 7 Metal coherence provided by the attraction between positive ions and an electron "gas" in a 2D close-packed array. Equivalent center distances are indicated by *dashed lines*. Such *close-packed structures* in 3D, such as face-centered cubic (*fcc*) and hexagonal close-packed (*hcp*) structures, in fact represent many metals: fcc, Al, Ni, Cu, Pd, Ag, Pt, and Au and hcp,Ti, Co, Zn, Zr, and Rh. Many other metals are also body-centered cubic (*bcc*): V, Cr, Fe, Nb, and Mo. It can be noted that these structures and elements having these structures are represented by vertical groups or columns in the periodic table of Fig. 5 in Chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements"

(electrons in a box), but this potential is not uniform within the box. Figure 7 illustrates this bonding model where the bonds from any ion are regarded as spherically distributed and act on as many neighbors as can be packed around that ion.

In alloy systems, most phases exhibit metallic bonding and the valence electrons are at least partially free. This applies to so-called intermetallic compounds in contrast to an intermediate solid solution alloy characteristic of noble metal alloys: Cu-Au and Cu-Ag. These alloys retain the fcc structure whether ordered or disordered. Hume-Rothery pointed out that some fundamental principles of alloy systems become apparent if the number of valence electrons per atom (or electron/atom ratio) is computed for each phase, and these have been referred to as Hume-Rothery compounds which possess wide ranges of solid solubility. We shall discuss these issues in more detail later. Suffice it to say that close packing is a key feature of metal and alloy structures (Fig. 8).

Van der Waals Bonding

Mutual interaction of atoms in separate molecules or during the interaction of molecules containing a hydrogen atom chemically connected with one of the atoms in the molecule (e.g., O or N) and some other atom of the same or neighboring molecule exhibits a so-called hydrogen bond. This can be intramolecular or intermolecular in nature, the latter characteristic of forces of mutual attraction of molecules. These intermolecular interactions also encumber the underlying theory of van der Waals forces which consist of *orientation* forces between molecular electric dipoles (the simplest being a molecular pair), mutual attraction as a consequence of *polarization*, and *dispersion* forces for nonpolar molecules.

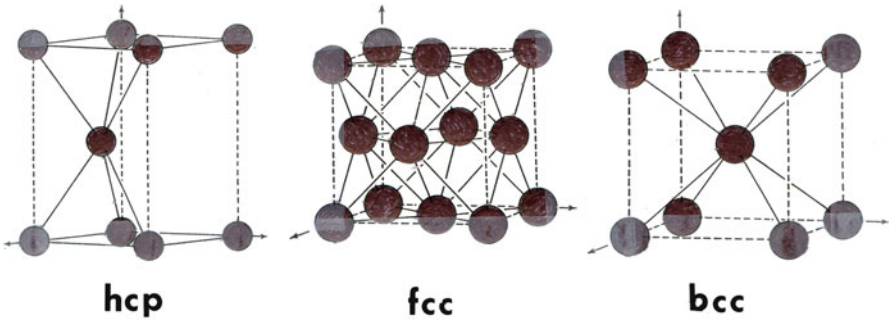


Fig. 8 Comparison of common metal structure unit cells

Dispersion forces correspond to the interaction of momentary electric dipoles which arise from electron spin. All molecules mutually attract regardless of their chemical nature. Saturated molecules or those incapable of forming a chemical bond also superimpose chemical forces of repulsion on the attractive van der Waals forces. The potential energy of these molecules may then be represented by the sum of these two forces:

$$U = U_o - \frac{A}{r^6} \tag{9}$$

where U_o represents the potential of repulsive forces, while the second negative term represents the attractive potential of van der Waals forces which is inversely proportional to the sixth power of the intermolecular distance, r. A in Eq. 9 assumes three different forms depending upon whether the interaction involves orientation forces, polarization forces, or dispersion forces. Dispersion forces are by far the largest contribution and more than an order of magnitude larger than the other two forces. Polarizability of the molecule measures the dipole induced in it by an electric field, and the larger the particle and more loosely bound its electronic structure, the greater the polarizability.

In the aggregate, the van der Waals force will be the potential energy of attraction times the area since energy is force/unit area. Consequently, for a very small graphite crystal, the van der Waals force will be small in contrast to a very large crystal, and large graphene layers will be much more difficult to separate. Van der Waals solids are therefore characterized by layer structures similar to simple graphite composed of covalently bonded carbon atoms in a hexagonal graphene lattice as shown in Fig. 4. Other common layer structure solids include talc and mica (Fig. 9) which, in contrast to graphite, are composed of more complex covalent layers bound by weaker van der Waals bonding. Such structures are characteristically *solid lubricants* since the layer separation allows contacting surfaces to slide. This phenomenon provides the

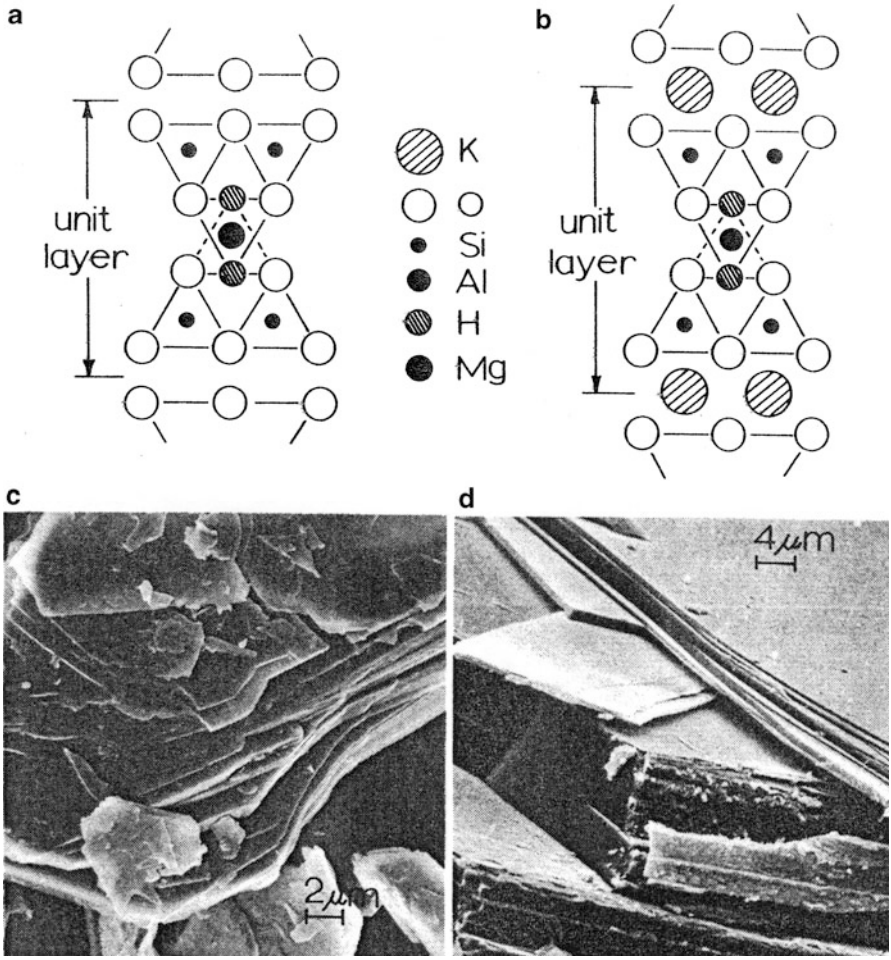


Fig. 9 Layer silicate structures. (**a**) Idealized plan edge view of monoclinic talc, $Mg_3(OH)_2Si_40_{10}$; (**b**) idealized plan edge view of muscovite mica, $KAI_2(OH)_2(Si_3A1)O_{10}$; (**c**) scanning electron microscope view of talc crystals; (**d**) scanning electron microscope view of mica crystals. Layer structures are apparent in (**c**, **d**)

practical basis for baby powder which is primarily talc, and the layer separation for talc particles feels smooth when applied to the skin. The separated platelets also provide a moisture guard for the skin. An interesting, contemporary example of van der Waals force is associated with the ability of the gecko to race across walls and ceilings even when the surfaces are irregular and dirty. K. Autumn and W. Hansen (an undergraduate student) at Lewis and Clark College in Portland, Oregon, around 2001 discovered this extraordinary

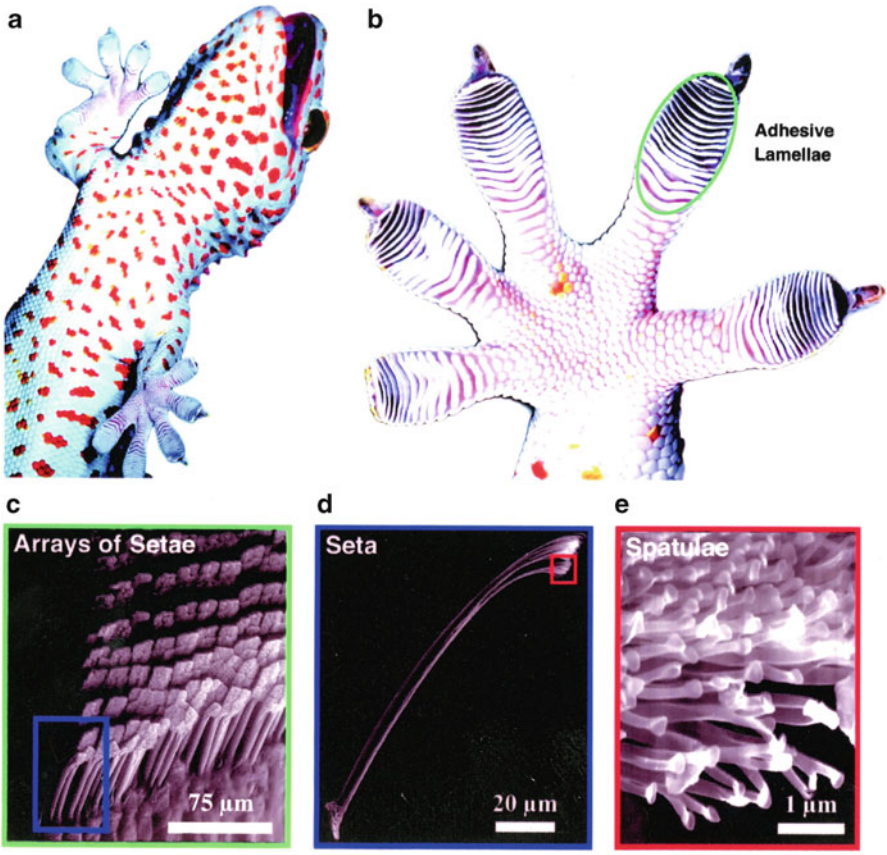


Fig. 10 Van der Waals force and gecko (**a**, **b**) mobility on walls and ceilings. (**c–e**) illustrate the arrays of nanofiber setae and their composition (From Hansen and Autumn (2005))

adhesive capability is due to the fact that the gecko feet possess adhesive lamellae or arrays of setae having spatulae at their ends measuring around 200 nm in diameter (Fig. 10) (Autumn et al 2002; Hansen and Autumn. 2005). These structures create a van der Waals force between the spatulae and the surface. As shown in Fig. 10c, there are roughly 10 spatulae/µm or about 10 million per square millimeter. This translates to an attractive van der Waals force of nearly 1 kg holding the gecko to the surface. Recognizing the potential for mimicking this response in creating a dry adhesive concept, closely grown arrays of carbon nanotubes have been demonstrated to exhibit similar effects (Autumn 2006). Similar biomimetic materials developments are treated in more detail in chapter "▶ Biomimetics and Biologically Inspired Materials."

References

Autumn K (2006) Gecko adhesion: structure, function, and applications. MRS Bull 32:473–478

Autumn K, Sitti M, Liang Y, Peattie A, Hansen W, Sponberg S, Kenny T, Fearing R, Israelachvili J, Full RJ (2002) Evidence for Van der Waals adhesion in gecko Setae. Proc Natl Acad Sci 99(19):12252–12256

Curzon AE, Mascall AJ (1965) The crystal structure of solid hydrogen and solid deuterium thin films. British J Appl Phys 16(9):1301–1309

Hansen WR, Autumn K (2005) Evidence for self-cleaning of gecko Setae. Proc Natl Acad Sci USA 102(2):385–389

Murr LE (2012) Chapter 1: Soot: structure, composition, and health effects. In: Paul MC (ed) Soot: sources formation and health effects. Nova Science, New York

Chemical Forces: Nanoparticles

Contents

Abstract

Atoms can aggregate or cluster in simple arrangements without forming a unit cell. Silica tetrahedral (SiO_3) and other tetrahedral clusters, double tetrahedral containing five atoms, pyramids containing five atoms, as well as octahedral or double pyramids containing six atoms are among the more fundamental clusters. Eight atoms form a simple cubic unit cell, while nine characterize a body-centered cubic cell, becoming more complex with face-centered cubic unit cells which can be rendered as icosahedrons of 12 atoms or 13 atoms with a body-centered or cluster-centered atom. These form building blocks for nanoparticles which can continue to add layers or shells forming layered or shell structures, even nanotubes. Many clusters and shell structures are represented by the platonic solids and their regular-face, convex polyhedra. Carbon clusters such as fullerene-based multilayer or multiconcentric clusters form unique nanoparticles. Variances of these structures form aggregates representing carbon-based soots and related nanoclusters.

Introduction

It is interesting that while atoms (or elements) of various kinds can bind together to form regular crystal structure units, some atomic species can also form regular, polyhedral structures which also reflect specific binding characteristics.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_8

Metal elements, metalloids, and covalent elements in particular can form such regular polyhedral and related structures, especially in the nano-regime, which can include nanoparticles ranging to hundreds of nanometers. While chapter "► Chemical Forces: Molecules" alluded to nanoparticle issues related to atom binding, nanoparticle specifics were not covered. Such particles can form around a central atom to form a polyhedral cluster. Correspondingly, this same polyhedron may form without the central atom to form a polyhedral cage. Such is the case for the C_{60} fullerene. The C_{60} hemisphere (C_{30}) can also extend to form a single-wall carbon nanotube. And just as multiwall carbon nanotubes can form, multiwall or multilayer fullerenes can also form, creating multi-shell, polyhedral nanoparticles and carbon nanoparticles.

In a similar way metal clusters can form polyhedral, multi-shell nanoparticles. In the case of such metal shell clusters, so-called superatoms are created. Indeed, even such superatoms can form as a polyhedral shell which itself can form a superatom in the regular atom in a crystal unit cell. In this chapter, these novel properties of atom clusters and shells will be described along with variances of polyhedral forms, including carbon nanotubes and relatively unstructured carbon (soot) nanoparticles.

Atom Clusters and Shells

Monocrystals and nanocrystal growth as implicit in ► Fig. 10b of chapter "Examples of Materials Science and Engineering in Antiquity" and ► Fig. 2 of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements" illustrate a range of novel phenomena including quantum mechanical phenomena, in contrast to bulk crystal phenomena. In the first case, these crystals can be considered to propagate polyhedral forms by simply adding atomic shells, while bulk crystals which grow from the melt (especially metals) can be considered to form by multiple nuclei propagating in a similar fashion but coalescing at some point of impingement to create large periodic atomic lattices in 3D. This requires long range translational periodicity which is not necessary for shell formation in growing polyhedra. This "shell periodicity" adds concentric layers of atoms in such a way that the overall symmetry of the resulting clusters is unchanged.

While bulk periodic crystals exhibit two, three, four, and sixfold rotational symmetry characteristic of the first three platonic solids (or polyhedral) shown in ► Fig. 1 of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements," small atom clusters can nucleate and grow with fivefold symmetry characteristic of the icosahedron and dodecahedron. In addition to the Platonic polyhedra, other polyhedral forms are possible such as the cuboctahedron and the truncated octahedron shown in ► Fig. 2 of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements." These structures represent lower energy configurations than the tetrahedron, octahedron, or cube. The faces of the Platonic polyhedra are triangles, squares, and pentagons. In 1611, Johannes Kepler introduced highly symmetric "modern" polyhedra having rhombic faces.

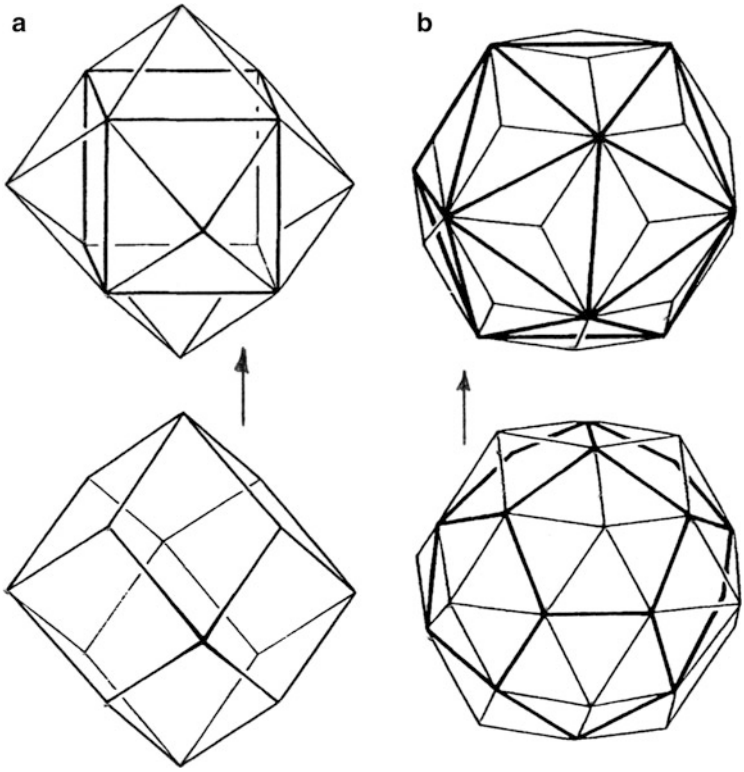


Fig. 1 Kepler polyhedra. (**a**) Rhombic dodecahedron in relation to a cube and octahedron. (**b**) Rhombic triacontahedron and its relationship with a decahedron and icosahedron

These included the rhombic dodecahedron with 12 rhombic faces and the rhombic triacontahedron with 30 faces. Figure 1 shows these polyhedra and their relationship to the cube and octahedron, and decahedron and icosahedron, respectively.

Low energy polyhedral structures favor close packing of the atoms composing them as well as small volume in relation to the polyhedron sphericity. Sphericity of a convex polyhedron such as the Platonic polyhedra and the cuboctahedron and truncated octahedron is defined as the ratio of the volume of the polyhedron and the volume of a sphere with the same surface area:

$$\text{Sphericity} = 6\pi^{\frac{1}{2}}\text{V}/\text{A}^{\frac{3}{2}} \tag{1}$$

and varies from 0 to 1 (perfect sphere) (Berman 1971). The cube has a volume of 1 and sphericity of 0.723 in contrast to the icosahedron with a volume of 2.18 and a sphericity of 0.91. The cuboctahedron has a volume of 2.36 and a sphericity of 0.86. The truncated octahedron has a volume of 11.3 and a sphericity of 0.82. Correspondingly, for very small polyhedral clusters, the icosahedron dominates followed by the cuboctahedron. Both can be related to close packing of sphere as discussed in

the context of metallic binding in ▶ Fig. 7 of chapter "Chemical Forces: Molecules." Figure 2 illustrates close-packed polyhedral derivatives for the face-centered cubic unit cell structure which characterizes bulk crystal structure periodicity for common metals as noted previously. These polyhedral derivatives show a central atom around which a polyhedral shell forms. Continual shell additions perpetuate the polyhedron. The clusters in Fig. 2 contain 13 atoms. But stable clusters can contain fewer atoms. For example, the SiO_3 tetrahedra forming glass structures in ▶ Fig. 3 of chapter "Electromagnetic Color and Color in Materials" can be regarded as clusters, and ideally these are close-packed clusters, with special bonds. For metal clusters, the concept of metallic binding illustrated in ▶ Fig. 7 of chapter "Chemical Forces: Molecules" is compromised for very small clusters with a small number of atoms. For example, clusters of lead atoms (where lead in bulk crystals has the fcc structure) have exhibited a propensity for 7 and 10 atoms (Duncan and Rouvray 1989), while sodium clusters exhibited 2, 3, 4, 6, 7, 8, 10, and 19 atoms (Bowlan et al. 2011). These numbers are often referred to as *electronic magic numbers*, and their existence has promoted the notion that clusters can be viewed as superatoms. This concept motivated the development of a so-called jellium model for atom clusters where the distribution of ion cores depicted in ▶ Fig. 7 of chapter "Chemical Forces: Molecules" is replaced by a uniform positive background where only valence electrons are treated explicitly.

The valence electrons affectively interact with the positive spherical charge distribution of the cluster, and energy levels can be calculated by solving an applicable form of the Schrodinger equation similar to that shown for the hydrogen molecule in ▶ Eq. 6 of chapter "Chemical Forces: Molecules". Figure 3a illustrates the jellium cluster model in contrast to atoms, represented by their energy levels, while Fig. 3b compares the valence energy levels for single atoms, small clusters, large clusters, and bulk metals. In principle, as clusters grow, the discrete energy levels begin to merge into bands. The electronic magic numbers of the atoms are 2, 10, 18, and 36 corresponding to the rare gases: He, Ne, Ar, and Kr (not shown in Fig. 3a) and 2, 18, and 40 for the clusters, corresponding to the electrons occupying the cluster energy levels and sublevels (Rao et al. 1999).

As the cluster grows and the free electron concept of ▶ Fig. 7 of chapter "Chemical Forces: Molecules" increases in prominence, cluster stability is determined by structure, and the magic numbers are referred to as *structural magic numbers* and refer to the total number of atoms depending upon the number of shells forming the polyhedral structure.

As observed in Fig. 2b, the cuboctahedron is a truncated cube, where truncation cuts a triangular face on the corners. The total number of atoms in a cuboctahedron shell structure or cluster is given by

$$N = \frac{10}{3}S^3 - 5S^2 + \frac{11}{3}S - 1 \tag{2}$$

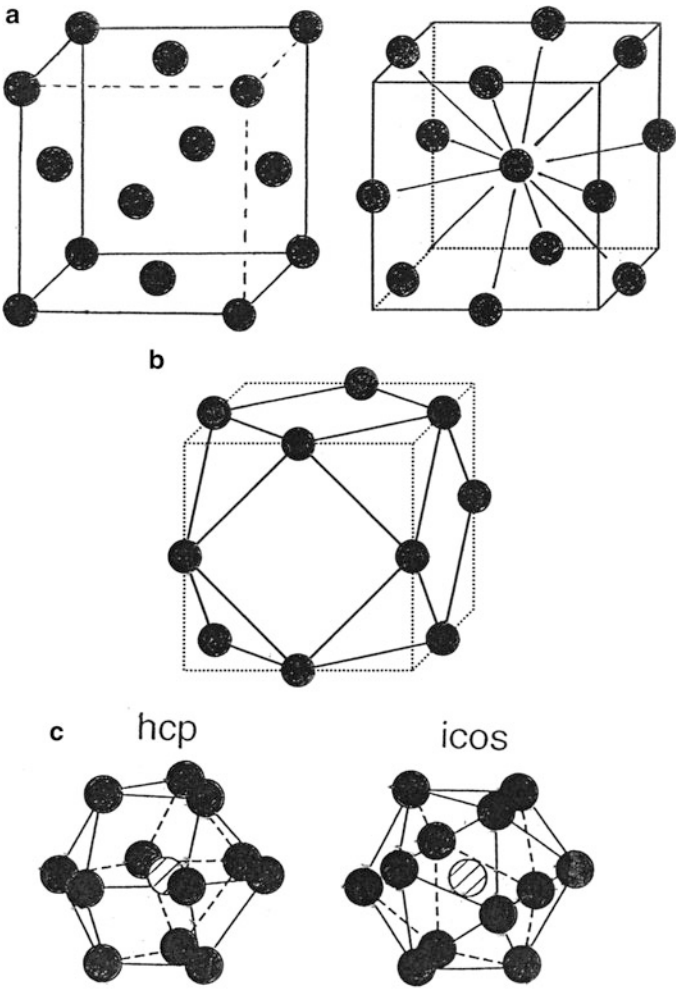


Fig. 2 Close-packed structure derivatives forming polyhedral shell structures. (**a**) Face-centered cubic unit cell and cell equivalent (with central atom). Atom in cell center is the first "shell." (**b**) Cuboctahedron formed from cell equivalent in (**a**). (**c**) Hexagonal close-packed (*hcp*) and icosahedral (*icos*) structures. These structures can be formed by selective stacking of close-packed planes or layers

where N is also the structural magic number corresponding to S polyhedron layers or shells. Here, as shown in Fig. 2b, the central atom is designated the first shell (S = 1). For the defining polyhedral shell, S = 2 and N = 13. Magic numbers for successive shells become 13, 55, 147, 309, 561, and 1415 as shown in Fig. 4. Figure 4 shows that the first full-shell cluster is formed by sequential stacking of close-packed layers starting with a rectangle and the central atom surrounded by six

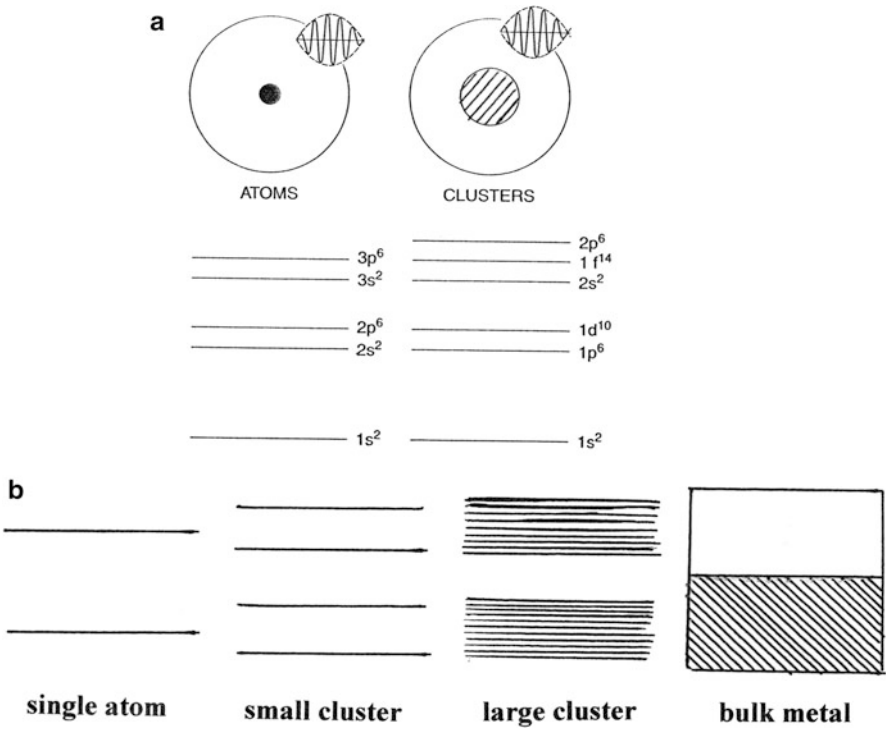


Fig. 3 Jellium cluster model (**a**) and change in valence energy band levels with cluster size (**b**) ((**a**) Adapted from Rao et al. (1999))

close-packed atoms next and a reversed triangle of close-packed atoms on top. For the third full-shell structure (S = 4) in Fig. 4, there are seven close-packed layers.

The number of atoms on the surface for any shell structure in Fig. 4 is given by the algorithm

$$\mathrm{N\,(surface)} = 10\mathrm{S}^2 - 2\mathrm{S} + 12 \tag{3}$$

which begins with the first full shell for S = 2. The corresponding diameter of each full-shell cuboctahedron nanoparticle is given by (2S−1) d where d is the distance between centers of nearest neighbor atoms measured across the fcc face diagonal in Fig. 2a, where $\mathrm{d} = \mathrm{a}/\sqrt{2}$; a is the unit cell lattice parameter. Starting with the first full shell, the particle diameters become $3\mathrm{a}/\sqrt{2}$, $5\mathrm{a}/\sqrt{2}$, $7\mathrm{a}/\sqrt{2}$, etc. As shown in Fig. 4, the cuboctahedron is constructed from close-packed layers, but the cut forming a square face shows a surface not close packed and energetically less favorable than the close-packed triangular face plane. However, the cuboctahedron has a small surface area and correspondingly large surface-to-volume ratio. This

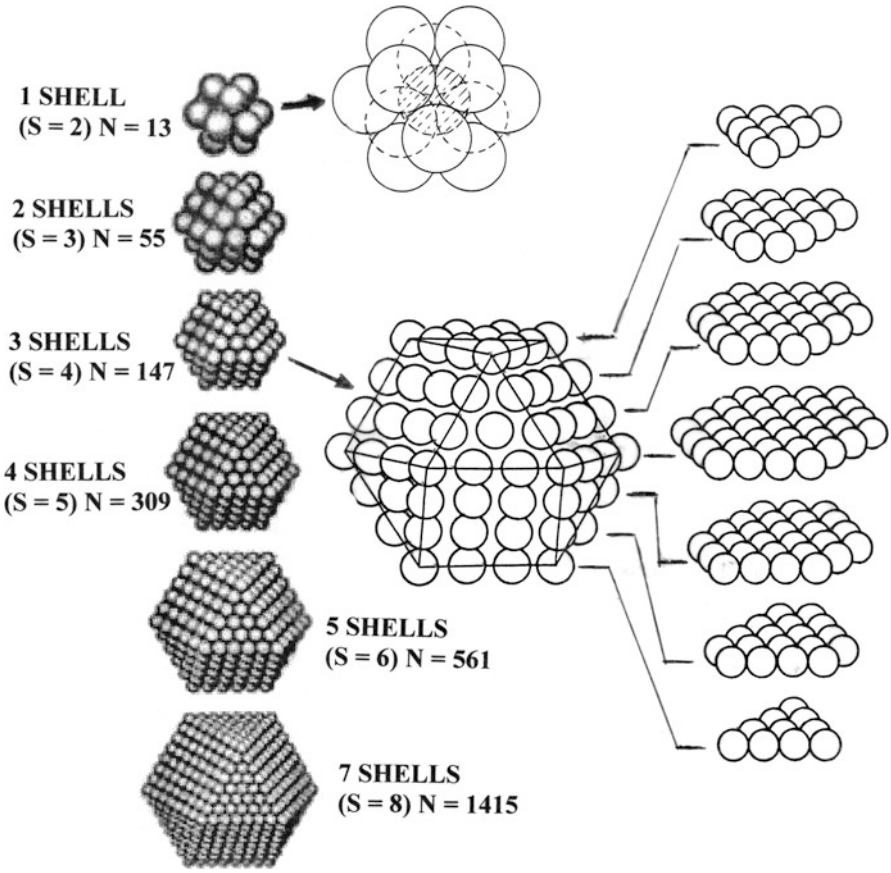


Fig. 4 Cuboctahedral cluster development, shell additions, and close packing of layers. Full-shell (S) clusters and structural magic numbers (N). Enlarged view is shown for N = 13 and 147. For N = 147, the layer structures are shown at *right*

means that the volume free energy component (E_V) is small relative to the surface free energy (E_S) since the total particle energy may be given by

$$\Delta G = \frac{4}{3}\pi r^3 \; E_V + 4\pi r^2 \, E_S \tag{4}$$

where ΔG is the total Gibbs free energy for a particle having a radius $r = (2S - 1)\, a/2\sqrt{2}$ and a is the fcc unit cell lattice parameter.

However, for a single full-shell cluster as shown in Figs. 2b and 4, the shell is considered the surface, and the single central core atom is the volume element, so the first term in Eq. 4 corresponds to 1 atom volume free energy. The second full

Table 1 FCC-cuboctahedral metal nanoparticle structural magic numbers (N) and related geometrical features for full-shell clusters

Full-shell number	Shell number (S)	Cluster diameter	Structural magic number (N)	Surface shell atoms (N(surface))	%[a]Surface
1	2	3d	13	12	92.3
2	3	5d	55	42	76.4
3	4	7d	147	92	62.6
4	5	9d	309	162	52.4
5	6	11d	561	252	44.9
6	7	13d	923	362	39.2
7	8	15d	1415	492	34.8
8	9	17d	2057	642	31.2
9	10	19d	2869	812	28.3
10	11	21d	3871	1002	25.9

[a]Percentage of atoms on the surface ((N/N (surface)) $\times 100$)

shell contains a total of 55 atoms with 42 composing the surface shell. Table 1 illustrates the geometrical features for a cuboctahedron particle.

A gold particle (a = 0.4 nm) with 10 full shells corresponding to S = 11 in Table 1 would have a diameter of around 6 nm and contain more than 3,800 gold atoms. Figure 5 shows a high-resolution TEM image for gold particles. The two smaller particles to the left have diameters near 6 nm. The lattice images correspond to gold (200) and (220) planes.

Icosahedral clusters have only close-packed surface faces or polygons, and in contrast to 6 square faces and 8 triangular faces for the cuboctahedron, it has 20 triangular faces, and with a relative volume of 2.18 and sphericity of 0.91 in contrast to 2.36 and 0.86, respectively, for the cuboctahedron, corresponding cluster sizes have lower energy for the icosahedron. Icosahedral clusters are referred to as nested Mackay icosahedra (Mackay 1962). As shown in Fig. 2c, this structure is similar to the cuboctahedron, and total number of atoms in Mackay icosahedron composed of S shells (or the structural magic number (N)) is identical to that for the cuboctahedron as represented by Eq. 2.

Table 1, correspondingly, will apply to Mackay icosahedral clusters. Figure 6a illustrates a 55-atom icosahedron and a portion of the next shell, while Fig. 6b shows a 2,869 atom icosahedron (Table 1).

The notion of icosahedral clusters as they manifest the most dense local packing of atoms is a fundamental concept for metal atom coordination and metal alloys, especially varied binary compositions. Figure 7 illustrates this concept for WAl_{12} from the work of R. C. Evans (1964) dealing with crystal chemistry which dates to 1939. But icosahedral quasicrystals began to appear as a wide range of cluster sizes in the mid-1980s, and more than 100 stable icosahedral quasicrystals have been identified, including a range of binary and ternary alloys (Tsai 1999; Tsai 2003). These aperiodic arrangements of atoms with no periodic crystalline unit cell

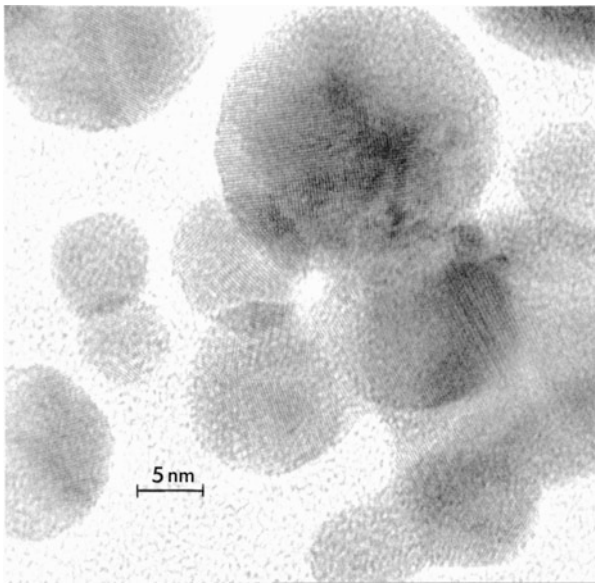


Fig. 5 High-resolution TEM image for gold nanoparticles

structures – incompatible with the translational symmetry of periodic crystal structures – undergo electronic transitions implicit in Fig. 3 as well as structural transformations to periodic crystal structures at very large sizes. As icosahedral quasicrystal clusters, metals and alloys can exhibit remarkable properties involving catalysis, unusual magnetic properties (even a net magnetic moment in small clusters composed of nonmagnetic atoms), and melting point suppression relating to the high surface-to-volume ratio. In addition, small icosahedral alloy clusters can exhibit unusual energetic phenomena such as segregation of the lowest surface energy element to the cluster surface and other atomic arrangements. These phenomena are illustrated for bimetallic particles in ▶ Fig. 32a, b of chapter "Electromagnetic Fundamentals" which show lowest energy substitutional isomers of noble metal atoms (Zhang and Fournier 2006). Lower energy Ag atoms aggregate to the surface in $Cu_{13}Ag_{42}$ and $Ag_{13}Cu_{42}$ where 12 of the 13 Ag atoms form an icosahedron at the polyhedron surface. In $Au_{13}Cu_{42}$ in ▶ Fig. 32b of chapter "Electromagnetic Fundamentals," the 13 Au atoms are on the surface, but they occupy different positions than Ag in $Ag_{13}Cu_{42}$. The corresponding (bulk) surface energies for these metal atoms are 1.675/m^2 for Cu, 1.415/m^2 for Au, and 1.145/m^2 for Ag (Samorjai 1994). These values simply provide a quantitative comparison.

In Fig. 8c, the Al-Li-Cu shows concentric structures of three types of icosahedral clusters: the Al-Mg-Si ternary class (Tsai 2008) in ▶ Fig. 32c of chapter "Electromagnetic Fundamentals," the Al-Li-Cu ternary class (representing Zn-Mg-Al class alloys in Fig. 8d), and the Cd-Yb binary class. In Fig. 8c the 55-atom Mackay icosahedron has a vacant center. The first shell is on Al/Si icosahedron, the second shell is an Mn icosahedron, and the third shell is an Al/Si icosidodecahedron. In Fig. 8d, the center of the first Al/Cu icosahedral shell is vacant, while the second

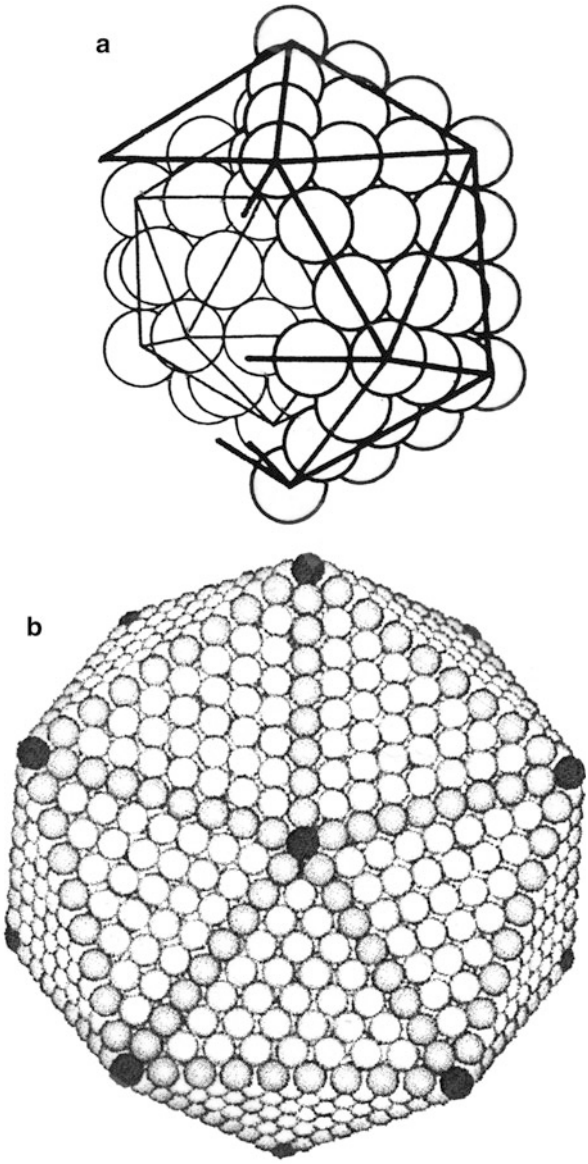


Fig. 6 Icosahedral clusters. (**a**) Closed-shell 55-atom icosahedron and a partial next layer (*heavy lines*). (**b**) Large icosahedral cluster: 2,869 atoms (N)

shell is an Li dodecahedron. The third shell is an Al/Cu icosahedron. In Fig. 8e, the center is a Cd tetrahedron, the first shell is a Cd dodecahedron, the second shell is a Yb icosahedron, and the third shell is a Cd icosidodecahedron.

This class is related to the Cd-Yb binary which has a body-centered cubic packing of identical icosahedral clusters (Tsai 2008).

Figures 7 and 8 provide a glimpse of the very unusual cluster and shell structures composing icosahedral nanoparticles. They demonstrate the evolution of atomic

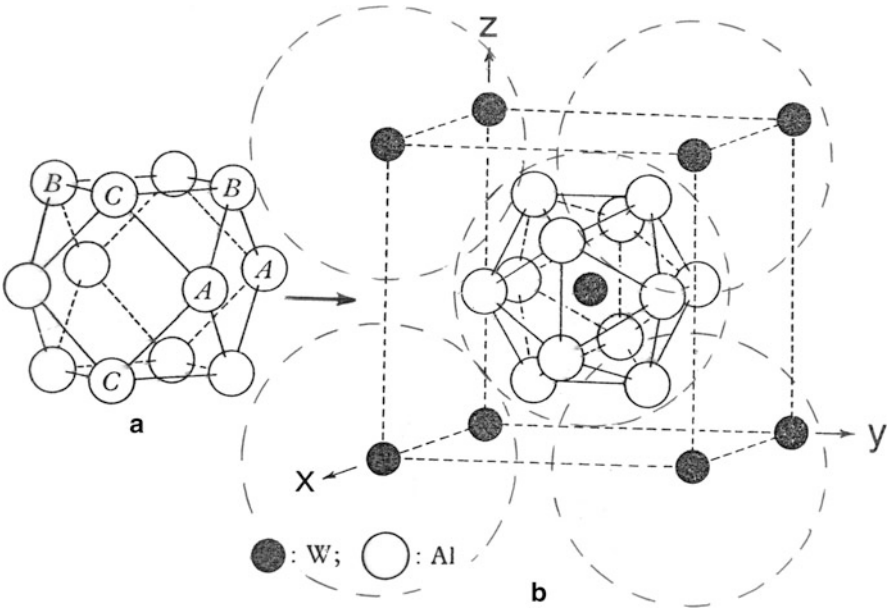


Fig. 7 (**a**) FCC derivative cuboctahedron coordinating close-packed atoms. (**b**) Cubic WAl_{12} structure unit. W atoms at the *corners* and cell center are surrounded by 12 Al atoms at the corners of a regular icosahedron, only one of which is shown. *Dotted circles* represent these superatom clusters at the cell vertices. Only the front face is represented. This icosahedron is derived from the cuboctahedron in (**a**) by displacing the atoms to reduce the distances A-A, B-B, C-C, etc. (Adapted from Evans (1964))

and molecular configurations defining quantum mechanical connectivity and the novelty of alloy design possibilities in contrast to bulk, periodic crystal structures which continue to dominate alloy metallurgy.

Carbon Shells and Clusters

Although we have discussed carbon nanotubes and black carbon from both a historical and contemporary perspective in chapter "▶ Examples of Materials Science and Engineering in Antiquity," these represent divergent structures: regular graphene shells forming nanotubes versus irregular, overlapping, curved graphene fragments and intercalated polycyclic aromatic hydrocarbon (PAH) isomers forming amorphous or turbostratic soot clusters (▶ Figs. 3,▶ 4, ▶ 5, ▶ 7 and ▶ 8 of chapter "Examples of Materials Science and Engineering in Antiquity"). Here we discuss regular, polyhedral shells and nested shells or clusters of carbon forming fullerenes, hyperfullerenes, and carbon onion structures. The original fullerene shell structure, C_{60}, characterized by the truncated icosahedron with 12 pentagonal faces, 20 hexagonal faces, and 60 vertices acting as the carbon atom sites was

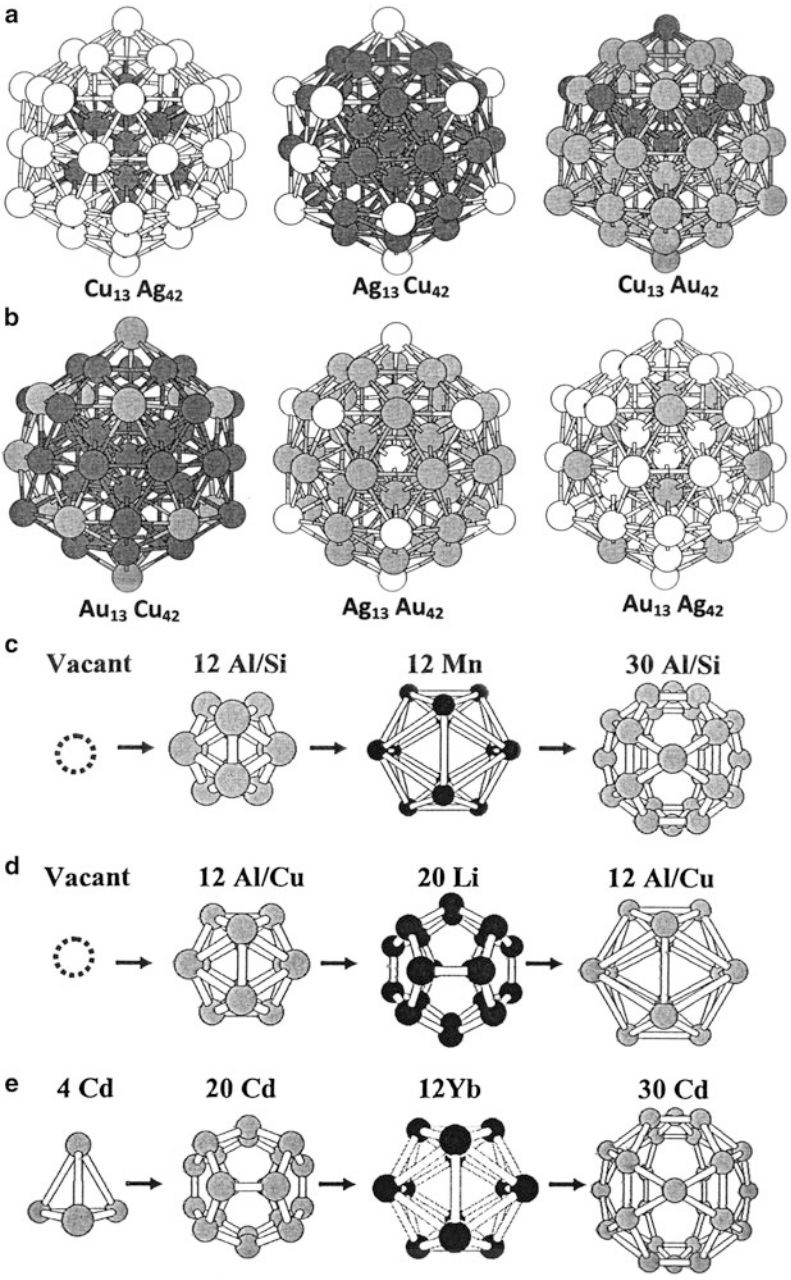


Fig. 8 Examples of bimetallic and trimetallic atom clusters. (**a**) and (**b**) show 55-atom (2-shell) Mackay icosahedra for bimetallic noble-metal clusters (adapted from Zhang and Fournier (2006)). (**c**–**e**) show concentric structures for three types of icosahedral clusters discussed in the text (Adapted from Tsai 2008)

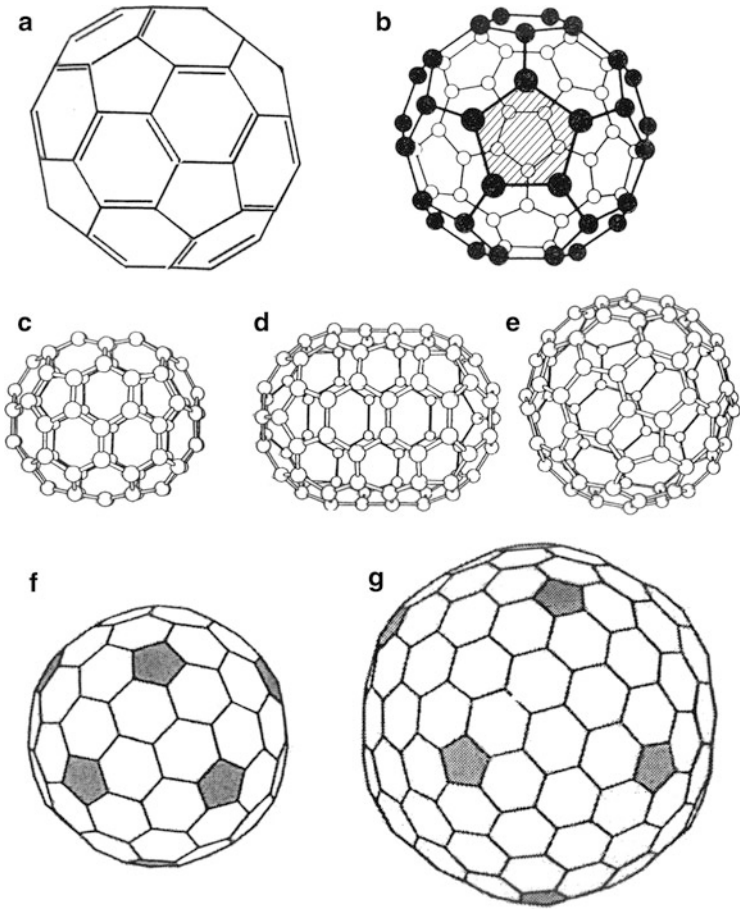


Fig. 9 Fullerene shell structures. (**a**) sp^2 icosahedral C_{60}. (**b**) Pentagon-centered rendering of the C_{60}-truncated icosahedrons. The shell diameter is 14.2 Å. (**c**) C_{70} icosahedral shell. (**d**) Elongated C_{80} shell. (**e**) Regular icosahedral C_{80} shell. (**f**) Icosahedral C_{140} shell. (**g**) Icosahedral C_{260} shell ((**c**–**e**) Adapted from Dresselhaus et al. (1996). (**f** and **g**) Adapted from Terrones and Terrones (2003))

discovered in 1985 by Kroto and colleagues and later named buckminsterfullerene (or buckyballs) in honor of R. Buckminster Fuller, the renowned architect of geodesic dome structures (Kroto et al. 1985). This structure discovery was awarded a Nobel Prize in 1996. Similar icosahedral shell structures and elongated shell structures extending to form varying diameters of carbon nanotubes also form from hemispherical shell caps which can close off immediately forming regular fullerenes, or extend before closing, even forming concentric and multiconcentric shell structures implicit in the simulated, ab initio-calculated shell structures shown in ► Fig. 4 of chapter "Examples of Materials Science and Engineering in Antiquity." Figure 9a, b shows the sp^2-truncated icosahedral carbon shell structure for C_{60},

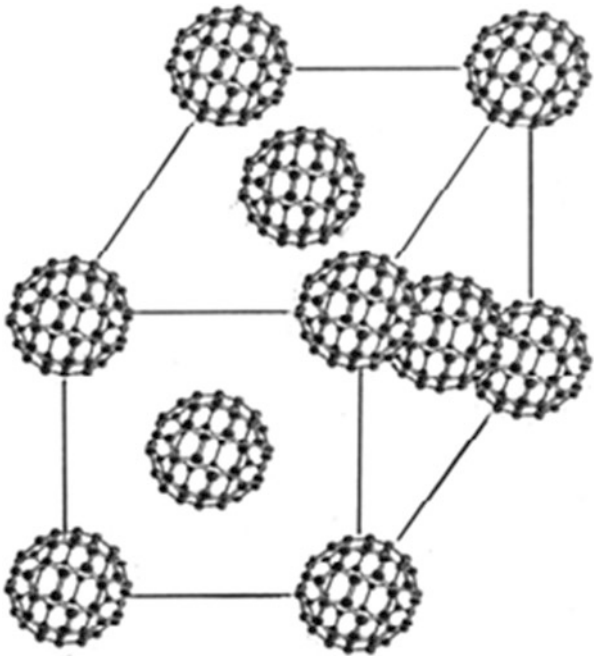


Fig. 10 FCC C_{60} (fullerite) crystal unit cell

while Fig. 9c shows a slightly distorted (elongated) truncated icosahedron for C_{70}, and Fig. 9d, e shows the elongated and regular icosahedral shell structures, respectively, for C_{80}. Figure 9f, g shows icosahedral shells for C_{140} and C_{260}. As noted earlier, regular icosahedral shells as well as closed, elongated shells require 12 pentagons (according to Euler's law) to create requisite curvature for the two associated hemispherical caps, each containing 6 pentagons. In principle, each pentagon contributes $\sim 30^\circ$ of curvature, with elongated portions separating the two closed ends containing only circumferential belts of hexagons.

The diameter of C_{60} is 7 Å. Around 1990, W. Kratschmer and colleagues (1990) found that solid crystals of C_{60} could be produced. These were yellow colored and as hard as diamond. They were composed of face-centered cubic unit cells of C_{60} acting as superatoms as shown in ▶ Fig. 34 of chapter "Electromagnetic Fundamentals." The unit cell lattice parameter was shown to be 14.3 Å, and this crystalline form of C_{60} has become known as *fullerite* (Fig. 10).

The formation of pentagonal- and hexagonal-based end caps from ab initio-based computer simulations by Lair et al. (2006, 2007, 2008) is reproduced in Fig. 11. These examples, referenced to the 2D graphene lattice shown in ▶ Fig. 5b of chapter "Examples of Materials Science and Engineering in Antiquity," illustrate energetically favorable nuclei for forming or growing regular or elongated closed icosahedral shell structures, including variously extended carbon nanotubes (CNTs). In fact, it has been shown that single, double, or possibly multiwall carbon nanotubes become energetically more stable as the tube lengths increase and

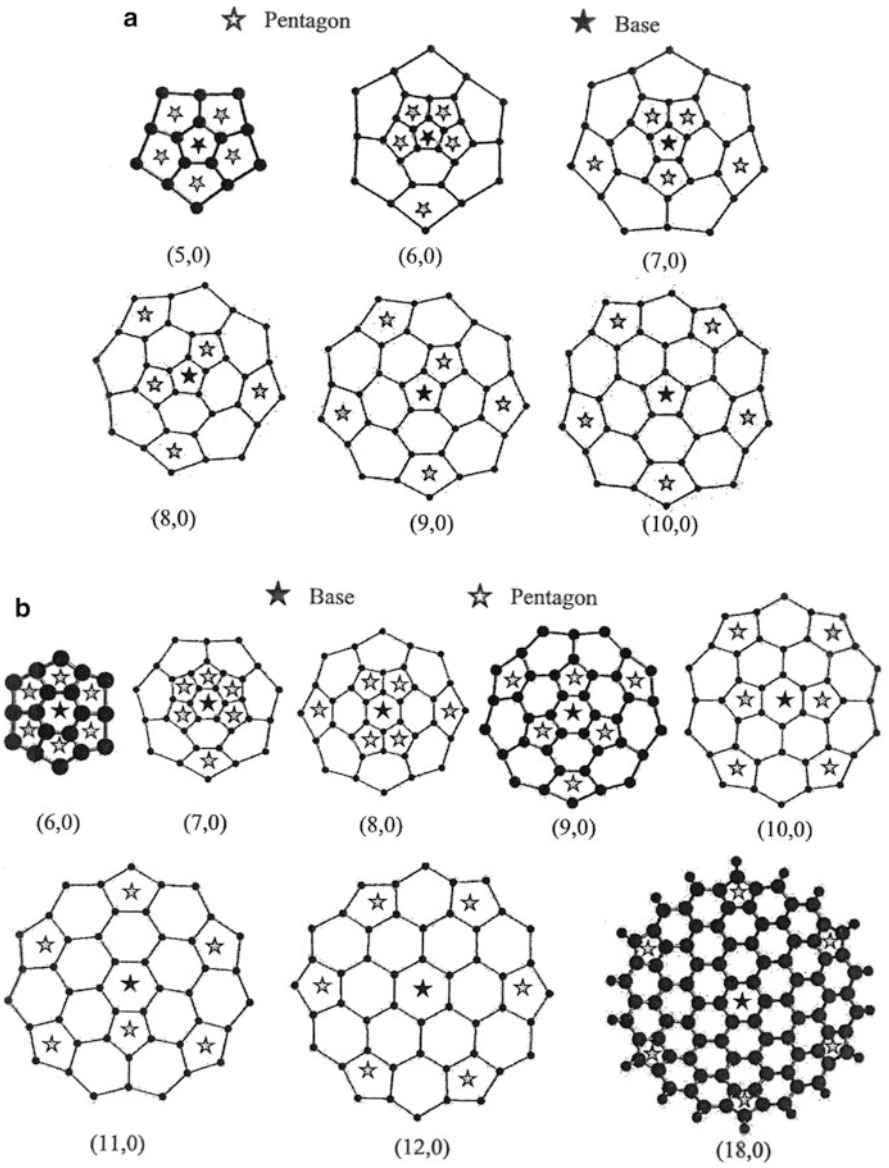


Fig. 11 Maps of pentagonal-based (**a**) and hexagonal-based (**b**) end caps ranging in circumference from 5 to 18 benzene rings corresponding to armchair CNT nuclei for $n = 5$ to $n = 10$; $m = 0$ (Adapted from Lair et al. (2006))

concentric tube diameters increase (Lair et al. 2008). Preformed end caps as shown in a few examples in Fig. 11 provide a nucleation template for controlled CNT growth. Slight thermodynamic or thermo-kinetic disturbances close the cap, forming fullerenes or slightly elongated fullerenes as shown in Fig. 9d.

These disturbances cause continuous, nested shell formation creating fullerene clusters as well as multiwall carbon nanotubes (MWCNTs). Guo et al. (1995) reported inherent MWCNT growth from carbon vapors implying that heterogeneous growth from an end cap is possible. In this regard, Lair et al. (2008) have shown that for multiwall shell configurations, the end caps and their combinations affected which tube combinations were more stable. The armchair/armchair (A/A) double-wall carbon nanotube formation shown in ▶ Fig. 4a of chapter "Examples of Materials Science and Engineering in Antiquity" has been shown to be the most energetically stable structure.

Carbon fibers, often produced by pyrolysis of organic (polymer) fibers such as nylon, etc., can have varying degrees of carbon (hexagonal) crystallinity date to the late 1800s when fibers were made from pitch. Tubular carbon filaments were first reported around 1976 using the TEM and further explored by Tibbetts (1984, 2001) in the 1980s using metal catalysts to grow them. Iijima (1991) obtained high-resolution TEM images showing multiwall carbon nanotube layer structures in 1991. The physical properties of carbon nanotubes have been described in considerable detail, including MWCNT elastic (or Young's) modulus on the order of 1.3–1.8 TPa (1,800 GPa), which is more than double the modulus for carbon fibers. Carbon fibers are, by contrast with MWCNTs, around 10 μm in diameter. Carbon filaments are a hundred times smaller in diameter and represent very large MWCNTs with irregular layer structure (Dresselhaus et al. 1988).

Certainly the exact growth mechanism for multi-shell carbon fullerenes and nanotubes is not well understood, and apparently subtle variations in the thermokinetic and chemical (hydrocarbon) environments in which they are grown have a significant effect as does the growth of CNTs and MWCNTs from catalytic nanoparticles such as Fe, Ni, and Co, rather than the pseudo-homogeneous nucleation on icosahedral end caps as illustrated in Fig. 11. Carbon atoms are added to the growing shells or tubes as hexagons or other polygons, and it is not known whether tube closures (for very long CNTs) are the same end caps as the nucleating end cap. Instabilities also create irregular, faceted fullerene clusters where the faceting is enhanced with growth, or the addition of shells, especially for icosahedral cluster symmetries. ▶ Figure 3 in chapter "Examples of Materials Science and Engineering in Antiquity" illustrates some of these structural features for multiconcentric fullerenes and MWCNTs grown in a carbon arc (by arc evaporation). Similar multiconcentric structures are observed in hydrocarbon flames such as methane or propane, the latter readily forming faceted fullerene clusters and MWCNTs as shown in Fig. 12. Figure 13a shows a high-resolution TEM image for a faceted, multi-shell fullerene cluster typical of those shown in Fig. 12. However, Ugarte (1992) has shown that these faceted, irregular fullerenes are energetically unstable and when heated to high temperatures will transform into lower energy quasi-spherical graphitic onion-like clusters shown in Fig. 13b. This transformation is illustrated ideally in Fig. 13c, d which shows 5-shell, C_n clusters, $n = 60$, 240, 540, 960, and 1,500, and assumes a C_{60}-truncated icosahedron at the center or first shell.

Fullerenes, CNTs, and soot are produced by arc discharge of graphite electrodes in inert atmospheres (such as He) (Yurdumakan et al. 2005), pyrolysis of

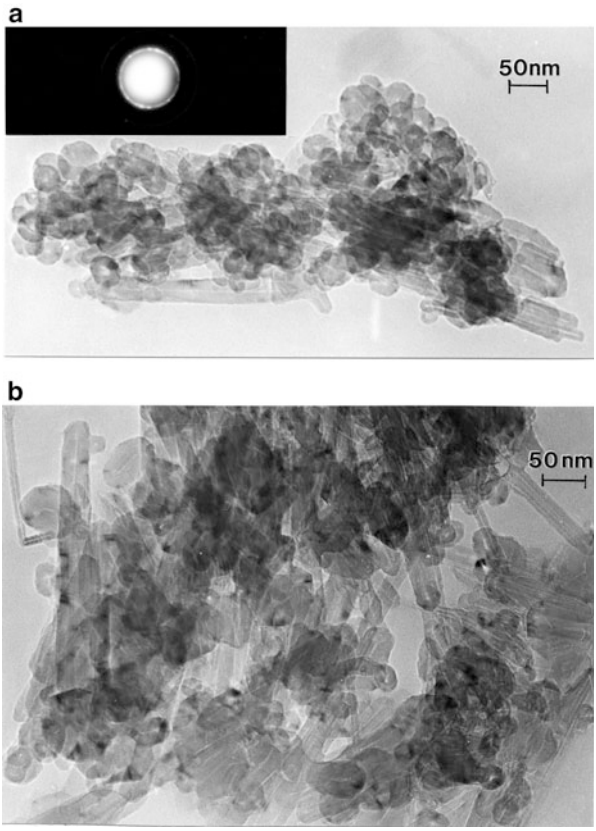


Fig. 12 Faceted, multiconcentric (hyper) fullerenes and MWCNTs collected from a propane (C_3H_8) flame and observed in the TEM. (**a**) Sample containing many multilayer fullerenes. Electron diffraction pattern *insert* shows graphite reflections, some spots on graphite diffraction rings indicating a degree of crystallinity. (**b**) Magnified view showing nanotubes and fullerenes

hydrocarbons over catalysts or catalytic chemical vapor deposition (Hamada et al. 1992), laser ablation, and electrolysis of graphite electrodes in molten salts (Chopra and Zettl 1998). In addition, and as shown in Figs. 2.3, 2.7, and 8.12, CNTs (especially MWCNTs) are produced in many combustion regimes, including candles, burning wood, and other flames. Figures 14 and 15 illustrate the structures typical for catalyst-grown and electric arc-grown MWCNTs, respectively. In contrast, Fig. 16 shows a wide range of fullerenes and MWCNTs in an aggregate collected above a natural gas (methane) kitchen stove burner. Note the contrast between the 10 nm MWCNT shown at A in Fig. 16 and the 40 nm MWCNT shown at B. Especially notable is the faceted MWCNT end for the very large diameter tube at B in Fig. 16, consistent with the pronounced faceting for fullerenes with increasing cluster size noted above.

Dense arrays of MWCNTs have also been grown by chemical vapor deposition into glass furnace tube walls and similar substrates as illustrated in Fig. 17. These arrays have been shown to emulate the van der Waals force regimes characteristic of the gecko foot shown in ▶ Fig. 10 of chapter "Chemical Forces: Molecules," and the functional resemblance is striking on comparison of Fig. 17

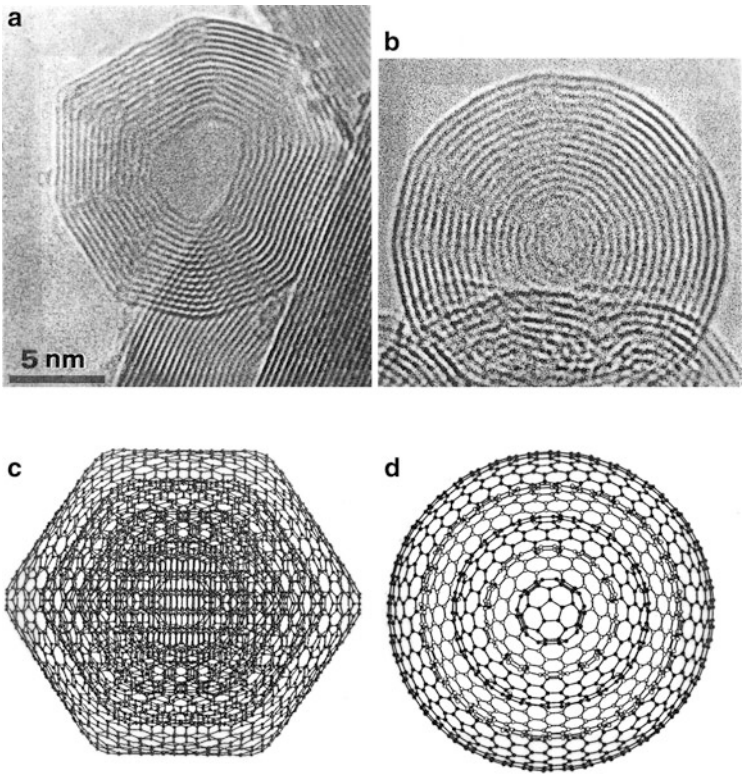


Fig. 13 Faceted-to-spherical carbon clusters. (**a**) High-resolution TEM image for faceted, irregular fullerene nanoparticle as in Fig. 12 (Adapted from Harris 1999). (**b**) High-resolution TEM image for quasi-spherical carbon onion (Adapted from Ugarte 1992). (**c** and **d**) How 5-shell faceted and quasi-spherical icosahedral-based structures. The rendering in (**d**) is a partial rendering

with ▶ Fig. 10 of chapter "Chemical Forces: Molecules." In addition, carbon nanotube arrays have also found uses in electron emission devices.

The MWCNTs shown in Fig. 16 represent tubes containing roughly 10 (Fig. 16a) to more than 50 concentric, cylindrical (graphene) shells (Fig. 16b). These concentric tube walls grow from continuously forming end caps as shown in ▶ Fig. 4a of chapter of "Examples of Materials Science and Engineering in Antiquity" and propagate one over the other, with the outer tubes lagging behind. This allows the inner tubes to close first, sometimes erratically and more often in the same spatial (3.4 Å) spaced sequence as the nucleating caps. Figure 18 shows simple hexagonal wire models for the carbon nanotube types and a double-wall structure for chiral tubes growing over an initial zigzag tube (Fig. 18d) and an armchair tube (Fig. 18e). It is believed that the overgrowth tubes as shown in Fig. 18d, e do not lag significantly behind the previous tube. As noted earlier, armchair/armchair sequences rather than zigzag/zigzag or chiral/chiral (as well as variations shown in Fig. 18d, e) are energetically more favorable. Hamada et al. (1992) demonstrated

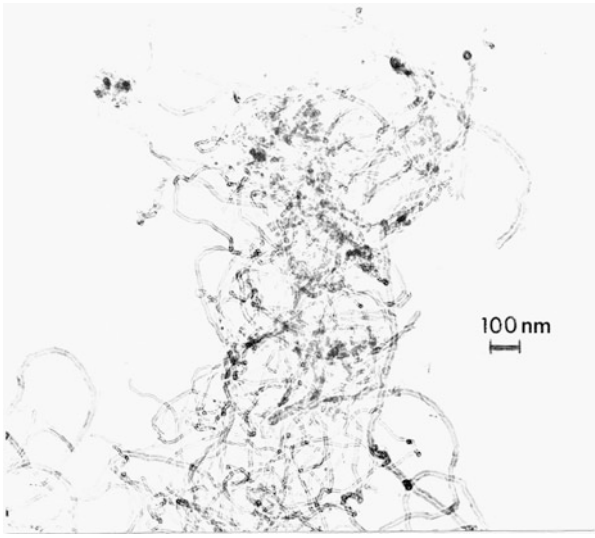


Fig. 14 MWCNTs grown from Ni catalytic nanoparticles observed in the TEM

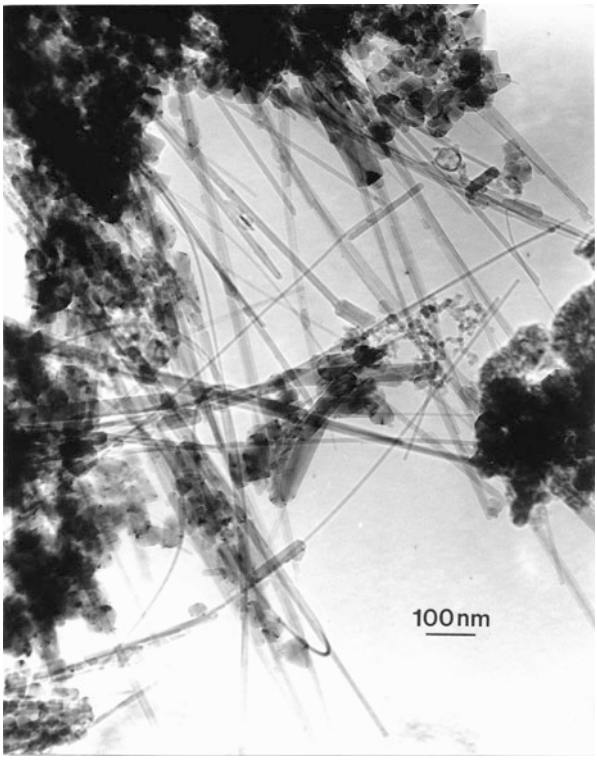


Fig. 15 Carbon-arc evaporation-grown MWCNTs observed in the TEM

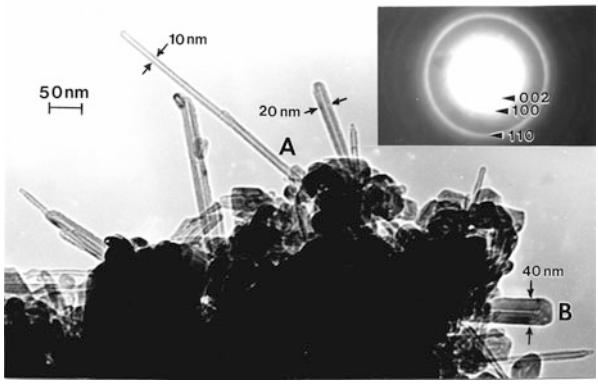


Fig. 16 Multiwall fullerene and carbon nanotube aggregate collected from a natural gas (methane) flame and observed in the TEM

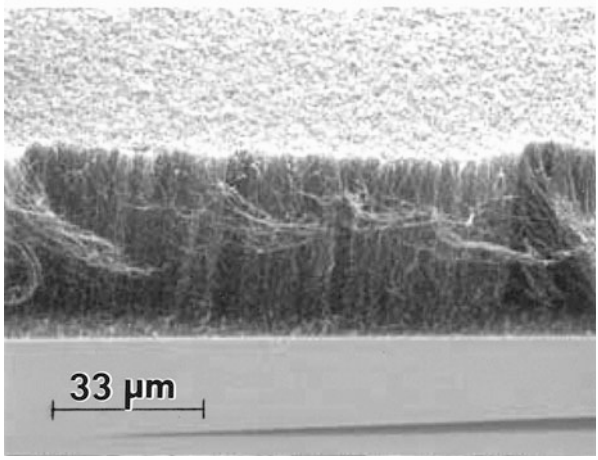


Fig. 17 Chemical vapor deposition-grown arrays of aligned carbon nanotubes

theoretically that the electronic properties of single-wall carbon nanotubes depended on their diameter and chirality. Zigzag nanotubes behave as semiconductors, while armchair tubes behave as metallic conductors. Exceptions occur when n-m is a multiple of 3 for zigzag tubes (see ▶ Fig. 5 in chapter "Examples of Materials Science and Engineering in Antiquity").

It is also frequently observed that especially large or multiwall carbon nanotubes such as that shown at B in Fig. 16 exhibit irregular or discontinuous layer structures. This feature is illustrated in Fig. 19 which shows an outer layer section of a large nanotube having more than 50 layers observed by high-resolution TEM. The insert in the lower right corner of the TEM image in Fig. 19 also shows for comparison the small nanotube segment at A in Fig. 16 and the corresponding, regular layer (graphene) structure.

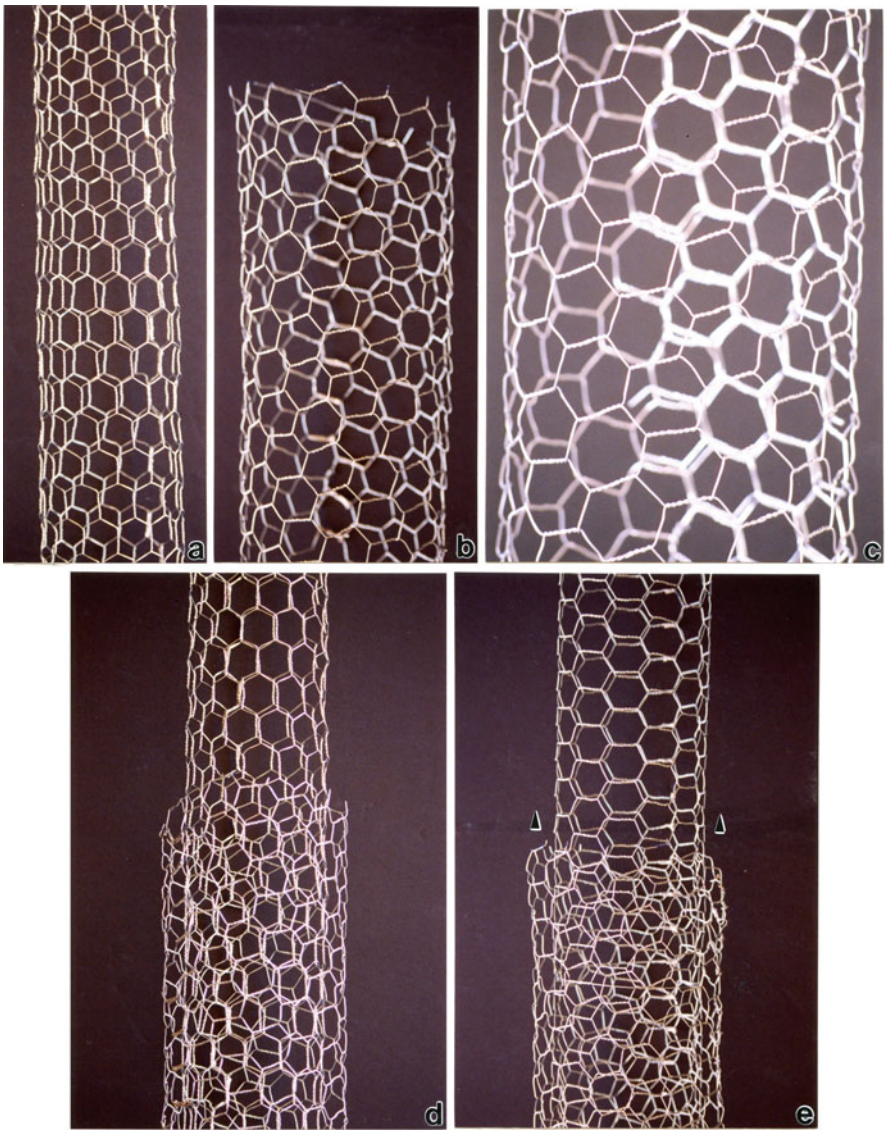


Fig. 18 Wire models for carbon nanotube types and concentric tube/wall growth. (**a**) Zigzag tube. (**b**) 13° chiral tube. (**c**) Magnified view of (**b**). (**d**) Chiral tube growing over zigzag tube. (**e**) Chiral tube growing over armchair tube. *Arrows* denote growth direction

The discontinuous and irregular carbon layer structure shown in Fig. 19 is even more irregular and amorphous-like for black carbon or soot nanoparticles produced in many regimes where fullerenes and carbon nanotubes are produced. Black carbon and soot nanoparticles are formed by curved graphene fragments

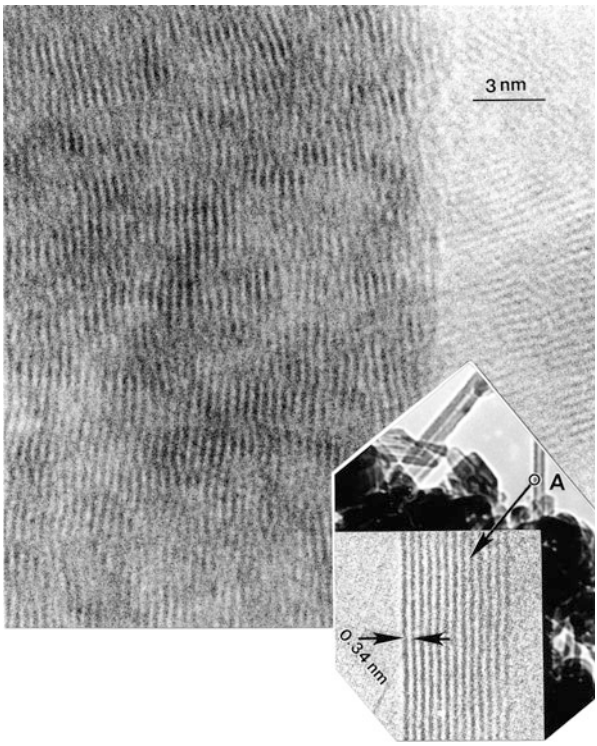


Fig. 19 High-resolution TEM image showing irregular, discontinuous carbon layer structure for large diameter MWCNT. The *insert* shows the corresponding structure for a small MWCNT having 13 cylindrical graphene layers

overlapping irregularly to form turbostratic clusters. In pyrolysis reactions involving various hydrocarbons, polycyclic aromatic hydrocarbon isomers can be intercalated within these curved graphene fragments as noted in chapter "► Examples of Materials Science and Engineering in Antiquity." In combustion regimes such as burning wood, combustion of diesel fuels, etc., a wide variety of PAH isomers having a range of molecular weights can be intercalated within the graphene fragments forming varying soot particle sizes.

Figure 20 illustrates a high-resolution TEM image of a black carbon nanoparticle cluster exhibiting turbostratic structure in contrast to the irregular carbon layer structure shown in Fig. 19. Figure 21 shows an evolutionary model for black carbon or soot particulate formation from curved fragments and intercalated PAH isomers, such as low molecular weight (MW = 128) naphthalene ($C_{10}H_8$). These nanoparticles usually form aggregates or chain aggregates illustrated for a number of soot clusters collected from various combustion regimes as noted in examples shown in Fig. 22. Commercial black carbon is produced in regulated pyrolysis processes which produce only low levels of low molecular weight PAHs such as naphthalene. Consequently, the nanoparticulates for good quality black carbon (BC) are relatively pure, turbostratic graphene fragments.

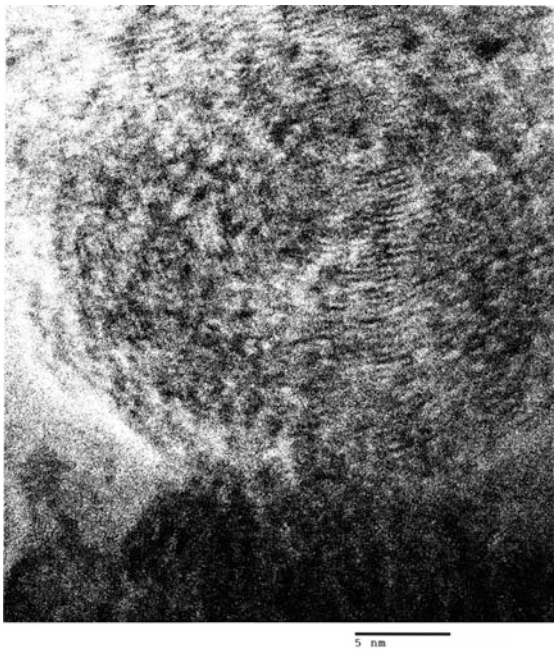


Fig. 20 High-resolution TEM image for black carbon nanoparticle

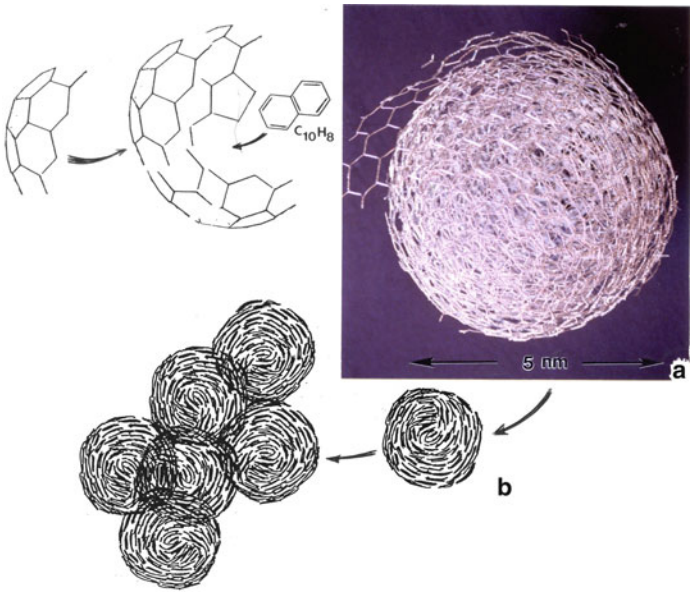


Fig. 21 Conceptual models for the formation of black carbon and soot nanoclusters and aggregated clusters. (**a**) Energetically favorable aggregation and intercalation of curved graphene fragments and PAH isomers and wire model for 5 nm soot particle. (**b**) Black carbon or soot nanoparticle aggregation (After Murr 2012)

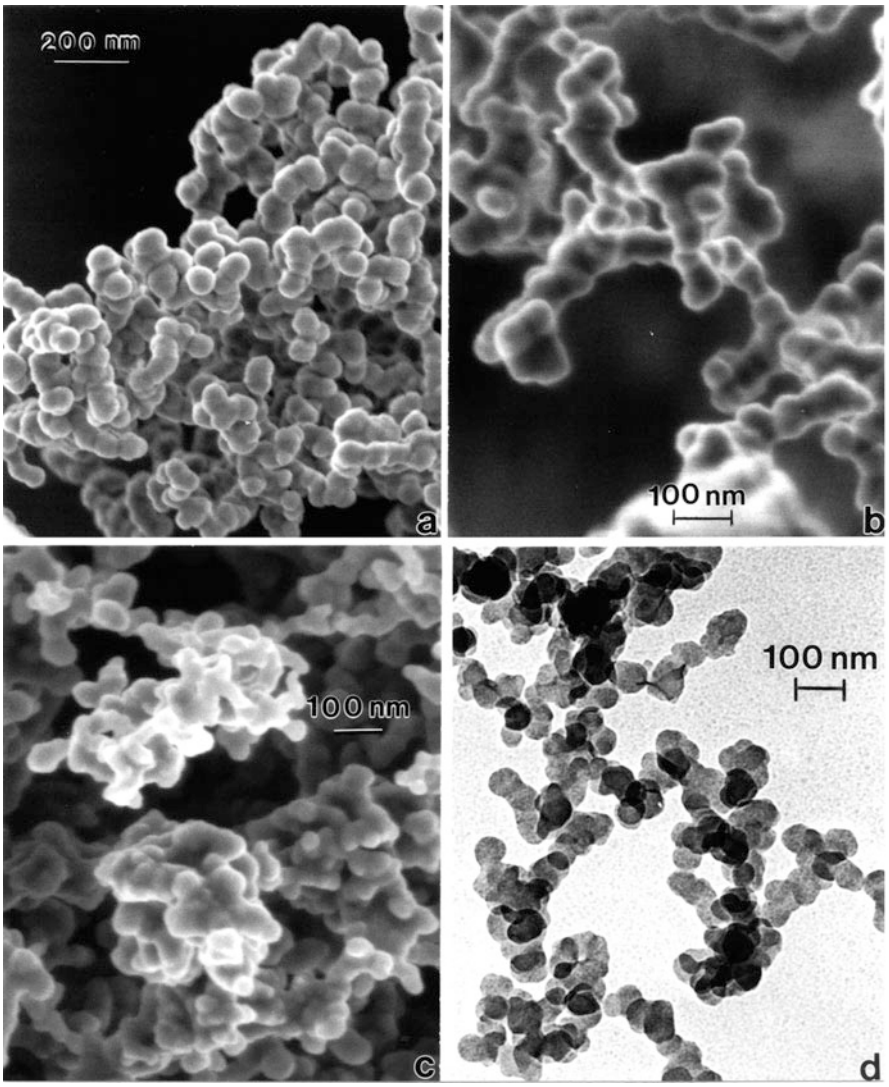


Fig. 22 Examples of soot nanoparticle aggregates. (**a**) Burning candle. (**b**) Wood burning. (**c**) Diesel truck exhaust. (**d**) Burning tire (Adapted from Murr 2012)

Nanotubes of Periodic Crystal Structures

As noted in chapter "▶ Examples of Materials Science and Engineering in Antiquity," nature produces mineral crystal nanotubes which are crystallographic (periodic) structures. Chrysotile asbestos, a nanotubular silicate as shown in ▶ Fig. 2 of chapter "Examples of Materials Science and Engineering in Antiquity," is an obvious example. Chrysotile asbestos nanotubes are also closed as grown, with the

end caps represented by periodic, hemispherical-like polyhedra such as cube octahedral or truncated cuboctahedra, with either a square or hexagonal base in contrast to the hexagonal or pentagonal base end forms shown for carbon fullerenes and nanotube in Fig. 11. An obvious nanotube material is hexagonal BN shown in ▶ Fig. 6 of chapter "Chemical Forces: Molecules" in contrast to hexagonal carbon (or graphite) in ▶ Fig. 18 of chapter "Electromagnetic Fundamentals." BN nanotubes, grown from a cuboctahedral hemisphere, produce layers separated ~3.33 Å, with B atoms interacting with H atoms in overlapping layers through a van der Waals force (Terrones and Terrones 2003). Unlike carbon nanotubes, BN nanotubes are insulators regardless of tube type, with a Young's modulus near the lower modulus for MWCNTs (around 1.3TPa) (Chopra and Zettl 1998).

While BN nanotubes consist of single sp^3-bonded BN layers as shown in ▶ Fig. 6 of chapter "Chemical Forces: Molecules," chrysotile asbestos and other crystal-like nanotubes are composed of more complex layer structures as shown in ▶ Fig. 2c of chapter "Examples of Materials Science and Engineering in Antiquity." Examples of these structures include metal sulfides such as MoS_2 and WS_2 which consist of the metal layer sandwiched between two sulfur layers. These triple layers are stacked together and separated in nanotube structures by 6.16 Å for MoS_2 and 6.18 Å for WS_2 (Terrones and Terrones 2003). These structures, formed from square-, hexagonal-, and octagonal-based end caps, have also been observed to form shell structures similar to fullerenes and more complex metal (nested shell) clusters illustrated in Fig. 8.

References

Berman M (1971) Regular-faced convex polyhedral. J Frankl Inst 291(5):329–352

Bowlan J, Liang A, de Haar WA (2011) How metallic are small sodium clusters. Phys Rev Lett 106(4):043401-1-4

Chopra NG, Zettl A (1998) Measurement of the elastic modulus of a multi-wall boron nitride nanotube. Solid State Commun 105:297–299

Dresselhaus MS, Dresselhaus G, Sugihara K, Spain IL, Goldberg HA (1988) Graphite fibers and filaments, vol 5, Springer series in materials science. Springer, Berlin

Dresselhaus MS, Dresselhaus G, Eklund PC (1996) Science of fullerenes and carbon nanotubes. Academic, New York

Duncan MA, Rouvray DU (1989) Microclusters. Sci Am 161:110–115

Evans RC (1964) An introduction to crystal chemistry, 2nd edn. Cambridge University Press, London

Guo T, Nikolaes P, Rinzler AG, Touranek D, Colbert DT, Swalley RE (1995) Self-assembly of tubular fullerenes. J Phys Chem 99(27):10694–10697

Hamada H, Sawada S, Oshiyama A (1992) New one-dimensional conductors: graphitic microtubules. Phys Rev Lett 68:1579–1581

Harris PJF (1999) Carbon nanotubes and related structures: new materials for the twenty-first century. Cambridge University Press, Cambridge

Iijima S (1991) Helical microtubules of graphitic carbon. Nature 354:56–58

Kratschmer W, Lamb LD, Fostiropoulos K, Huffman DR (1990) Solid C60. A new form of carbon. Nature 347:354–355

Kroto HW, Heath JR, Obrien SC, Curl RF, Swalley RE (1985) C_{60}: buckminsterfullerene. Nature 328:162–163

Lair SL, Herndon WC, Murr LE, Quinones SA (2006) End cap nucleation of carbon nanotubes. Carbon 44:447–455
Lair SL, Herndon WC, Murr LE (2007) Energetic trends of single-walled carbon nanotube ab initio calculations. J Mater Sci 42(5):1819–1827
Lair SL, Herdon WC, Murr LE (2008) Stability comparison of simulated double walled carbon nanotube structures. Carbon 46:2083–2095
Mackay AL (1962) A dense non-crystallographic packing of equals. Acta Crystallogr 15:916–918
Murr LE (2012) Soot:structure, composition and health effects, Chap. 1 in Soot: sources, formation and health effects, Paul MC (ed) Nova Science Publishers, New York
Rao BK, Khanna SN, Jena P (1999) Designing new materials using atomic clusters. J Clust Sci 10:477–485
Samorjai A (1994) Introduction to surface chemistry and catalysis. Wiley, New York
Terrones H, Terrones M (2003) Curved nanostructured materials. New J Phys 5:1–34
Tibbetts GG (1984) Why are carbon filaments tubular? J Cryst Growth 66:632–638
Tibbetts GG (2001) Vapor grown carbon fiber research and applications, achievements, and barriers. In: Biro LP, Bernardo CA, Tibbetts GG, Cambin P (eds) Carbon filaments and nanotubes: common origins, differing applications. Springer, New York
Tsai AP (1999) Metallurgu of quasicrystals, In: Stadnik ZM (ed) Physical properties of quasicrystals. Springer, Berlin
Tsai AP (2003) Back to the future-an account discovery of stable quasicrystals. Acc Chem Res 36 (1):31–38
Tsai AP (2008) Icosahedral clusters, icosahedral order and stability of quasicrystals-a view of metallurgy. Sci Technol Adv Mater 9:1–20
Ugarte D (1992) Curling and closure of graphitic networks under electron-beam irradiation. Nature 359:707–709
Yurdumakan B, Raravikar NR, Ajayan PM, Dhinojwala A (2005) Synthetic gecko toot-hairs from multiwalled carbon nanotubes. Chem Commons 30:3799–3801
Zhang M, Fournier R (2006) Structure of 55-atom bimetallic clusters. J Mol Struct (THEOCHEM) 762:49–56

Part IV

Crystal Structures and Crystal Chemistry

Crystallography Principles

Contents

Abstract

This chapter presents a short review or overview of point and space group symmetries and notations, and lattice systems and crystal structures and structure types.

Introduction

Regular, symmetrical lattices can be described in terms of 2D or 3D constructs. Graphene, as illustrated as a practical 2D construct in ▶ Fig. 5 of chapter "Examples of Materials Science and Engineering in Antiquity", represents the more classical 2D construct, while the graphite 3D lattice can be viewed as a corresponding 3D construct composed of overlapping 2D lattice arrays. The 14 Bravais lattices representing 7 crystal systems provide the basis for crystal structures, while the complete realm of 3D crystal structures is represented by 230 space group (S.G.) designations.

Lattice and Crystal Systems and Their Formalized Notations

The concept of a 2D point lattice has already been explored in ▶ Fig. 5 of chapter "Examples of Materials Science and Engineering in Antiquity" as it applies to hexagonal sheets of carbon, or graphene, and the characterization of carbon nanotubes and nanotube types. As observed especially in ▶ Fig. 5b of chapter

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_9

"Examples of Materials Science and Engineering in Antiquity", the 2D graphene lattice represents a repeating or periodic array of points. As noted in ▶ Fig. 5b of chapter "Examples of Materials Science and Engineering in Antiquity", any point relative to the origin at (n, m) = (0, 0) is related to any other point by a set of vectors

$$\mathbf{C} = n\mathbf{a}_1 + \mathbf{m}\mathbf{a}_2 \tag{1}$$

where |**C**| in ▶ Fig. 5 aof chapter "Examples of Materials Science and Engineering in Antiquity" represents the nanotube circumference and $\mathbf{a_1}$ and $\mathbf{a_2}$ are equivalent vectors having magnitudes of $|\boldsymbol{a}_1| = |\boldsymbol{a}_2| = 2.46\text{Å}$ for the graphene lattice; where the lattice points are replaced by carbon atoms, and $\mathbf{a}_1$ and $\mathbf{a}_2$, the fundamental translation vectors serve as reference axes for graphene, which can be considered a 2D crystal.

Alternatively, a 2D lattice can be formed by identifying unit cells and translating these cells parallel to their two edges. However, this operation can also be achieved by appropriate rotations of the unit cell around any point of the unit cell, especially rotations about an axis through a point perpendicular to the plane of the lattice. Symmetry operations include reflection (which reflects the unit cell across a reflection plane), rotation about on axis, inversion or change of sign of the coordinates of each point with respect to the center of symmetry, and improper rotation (rotation about an axis followed by an inversion). The arrangement of points creating a periodic lattice can be represented by so-called symmetry elements, and a collection of such elements at any point is called a point group. The unit cell translation combined with the symmetries at a point (or the point group) defines what is referred to as a space group. A general 2D lattice is defined by whether the edges (or unit cell vectors) are equal or not and the angle between them, creating four such axial systems (or unit cells) in 2D: square, parallelogram, rectangle, or rhombus. Lattices resulting from the translation of these unit cells have at least 1-fold and 2-fold symmetries; more three-, four-, or sixfold axes are involved. All but the parallelogram have mirror planes (Buerger 1960).

These 2D lattices can be sketched to create 3D space lattices just as graphene is stacked to create graphite (▶ Fig. 4 of chapter "Chemical Forces: Molecules") or other atoms in close-packed lattice planes are stacked as implicit in ▶ Fig. 4 of chapter "Chemical Forces: Nanoparticles". The parallelogram is an appropriate example as a starting point since the 3D space lattice unit cell it creates will have vectors $\mathbf{a_1}$, $\mathbf{a_2}$, and $\mathbf{a_3}$ and corresponding interaxial angles α, β, and γ all unequal. Figure 1 illustrates this construction periodicity. Figure 1a shows a general 2D lattice, while Fig. 1b illustrates the corresponding 3D lattice. Figure 1c shows the 3D lattice unit cell which is defined as the triclinic system. Note that the 2D lattice in Fig. 1a can either be stacked horizontally or vertically. In Fig. 1c, the unit cell vectors $\mathbf{a_1}$, $\mathbf{a_2}$, and $\mathbf{a_3}$ are replaced using the notation **a**, **b**, and **c**, respectively, and referred to as lattice parameters. When these serve as reference axes for a crystal unit cell, they provide the unit cell dimensions. Consequently, when Fig. 1c corresponds to a crystal unit cell, and all possible combinations of equality and inequality in crystal axes a, b, and c (the scalar values of **a**, **b**, and **c**) and interaxial angles,

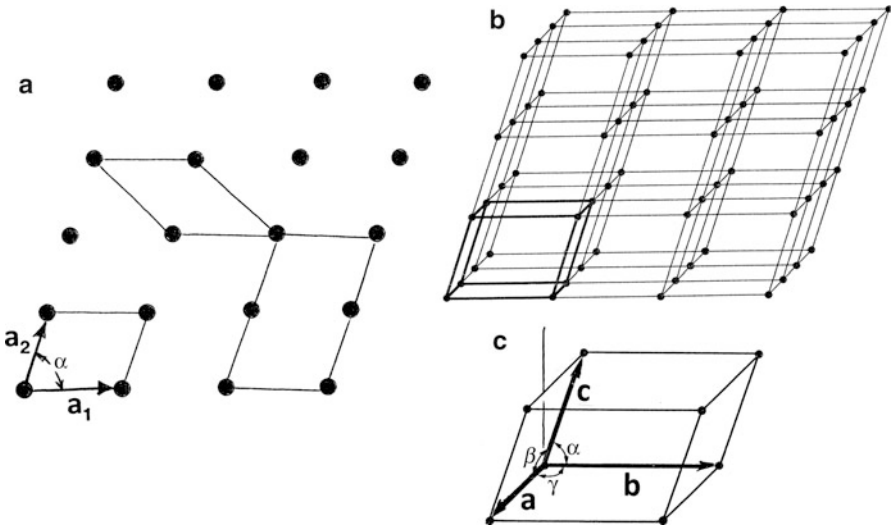


Fig. 1 Space lattices and unit cells. (**a**) 2D space lattice and unit cells. (**b**) 3D space lattice. The lattice unit cell is shown *bold*. (**c**) Details of the unit cell in (**b**) showing lattice parameters (axes) and interaxial angles

α, β, and γ, are counted, it is found that only 14 kinds of unique lattices are possible. This was first demonstrated by A. Bravais around 1850, and consequently these 14 lattice types have become known as *Bravais lattices* (Bravais 1850). Each of these 14 Bravais lattices has been shown to exhibit some convenient set of axes, each with corresponding and specified lattice parameters and interaxial angles, and a total of 7 crystal systems are used in crystallography. These are illustrated in Fig. 2 as crystal unit cells. In the 3D space lattice representing the scheme of repetition in a crystal, corresponding point groups are sets of symmetry operations that, like the 2D lattice, lack a central point or fixed point of reference, while crystal faces and directions are moved to replicating positions. Each crystal system represented in Fig. 2 has corresponding sets of point groups, and there are a total of 32 point groups for the 7 crystal systems. Each lattice system has a lattice symbol in the so-called Hermann-Mauguin notation.

The ways macroscopic and microscopic symmetry elements can be periodically arranged in space are represented by *space groups* which add centering information to point groups. Consequently, point groups and space groups are related, and like point groups, there are a specific number of space groups associated with each crystal system. There are a total of 230 space groups in the 7 crystal systems. Table 1 summarizes these crystallographic features as well as those implicit in Fig. 2.

It might be noted that all of the unique combinations of point groups associated with 2D lattices illustrated in Fig. 1a involving the possible rotations and mirror (m) planes result in 10 point groups consistent with 2D periodicity and are possible. Their designations for rotations are 1, 2, 3, 4, 6, and for mirror planes: 1 m, 2 mm

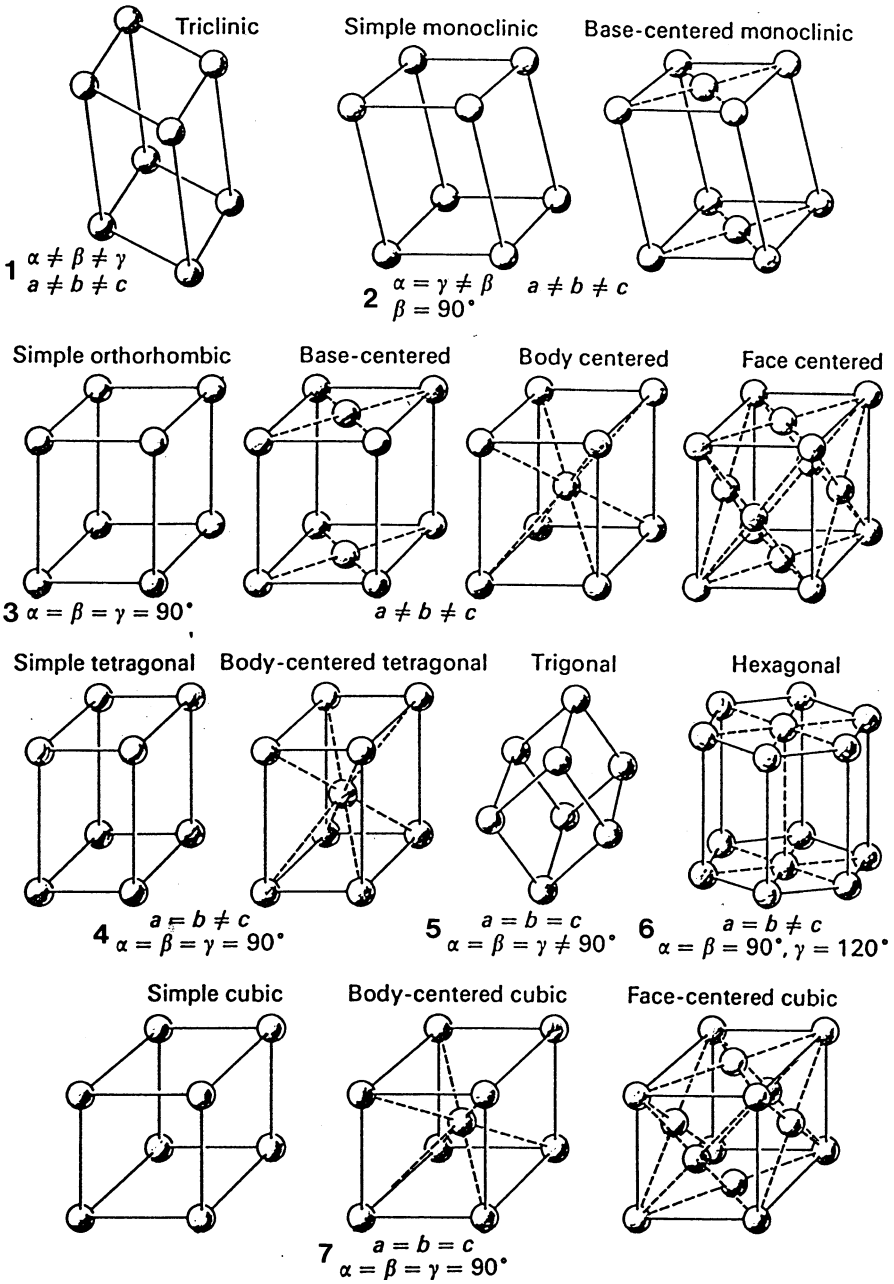


Fig. 2 The 14 Bravais (space) lattices and 7 crystal systems represented by 1 to 7. Lattice parameters (a, b, c) and interaxial angles (α, β, γ), referenced to Fig. 1c

(or mm), 3 m, 4 mm (or 4 m), and 6 mm (or 6 m), where 2 mm is written as mm because a twofold axis always occurs at the interaction of two perpendicular mirrors, etc. Consequently, 3D periodicity combines all possible symmetries in the 230 unique space groups, and any 3D (periodic) crystal must be one of these 230 space groups, designated S.G., either the space group number (1–230) or a space group designation (S.G.) which begins with the lattice symbol shown in Table 1, and includes the point group symmetry operations also shown in Table 1. Figure 3 shows some common crystal structures. Note that copper and iron represent fundamental crystal (or lattice) systems (Fig. 2) while NaCl and CsCl represent modifications of these systems. Notice that the NaCl structure represents two interpenetrating face-centered cubic structures: fcc Na and fcc Cl. The details of the symmetry elements and space group designations can be found in the International Tables for X-ray Crystallography (1952–1960) and the classic book by Buerger (1960).

It might be noted in correction with the Hermann-Maguin notation that the point group is simply a brief list of symmetry elements associated with the three crystallographic directions (**a**, **b**, **c** in Fig. 1c). The macroscopic symmetry elements in this notation as represented in Table 1 are as follows: 1-, 2-, 3-, 4-, and 6-fold rotation axes are denoted 1, 2, 3, 4, and 6; the symmetry plane is denoted m (for mirror plane) as indicated previously; axes of rotation-inversion are denoted 1, 2, 3, 4, and 6; and the center of symmetry is designed $\overline{1}$ (which is equivalent to the 1-fold rotation-inversion). These notations are implicit in the space group designation for cubic crystal unit cells in Fig. 3.

In addition to space group (S.G.) designations, crystal structures are often also designated by a common structure type referred to as the *Strukturbericht* designation. Metals and other elements are designated by A with a number: fcc structures are designed Al; bcc, A2; hcp, A3; diamond cubic (dc), A4; tetragonal, A5/A6; rhombohedral, A7; and hexagonal (not hcp or cph), A8/A9. The crystal structures and lattice parameters for the elements are summarized in the periodic table in Fig. 4. It can be observed in the periodic table in Fig. 4 that there is often a clear relationship between crystal structure and the position of elements in the table. For example, elements in the same group tend to have the same structure, e.g., Cu, Ag, Au, etc. It is also observed in Fig. 4 that many elements have different crystal structures or *allotropic forms*, a phenomenon called *polymorphism*. This often occurs at different temperatures and/or pressures. Allotropic forms are designated as crystal phases (or forms): α-Fe (bcc phase), γ-Fe (fcc phase), and δ-Fe (bct phase). The room temperature α-Fe (ferrite) transforms to γ-Fe (austenite) at 910 °C and becomes δ-Fe (ferrite) above 1,400 °C. Cobalt exists as hcp (ϵ) and fcc (δ). The most extraordinary polymorphic metal is plutonium which has six phases or crystal structures: α, β, γ, δ, δ^1, and ϵ. Plutonium metallurgy will be discussed later.

Other common Strukturbericht designations include the NaCl structure (Bl), CaF_2 structure (Cl), and other composition types. Table 2 summarizes the Strukturbericht designations corresponding to crystal compositions. In contrast to

Table 1 Summary matrix for crystal systems

Crystal system	Point group symmetries	# of point groups	# of space groups	Space group # (S.G.)	Bravais lattices	Lattice system	Lattice symbol
Triclinic	None	2	2	1, 2	1	Triclinic	P (simple)
Monoclinic	1 2-fold axis of rotation or 1 mirror plane	3	13	3–15	2	Monoclinic	P (simple)
							C (bc)[a]
Orthorhombic	3 2-fold axes of rotation or 1 2-fold axis of rotation plus 2 mirror planes	3	59	16–74	4	Orthorhombic	P (simple)
							I (bc)[a]
							C (base)[b]
							F(fc)[c]
Tetragonal	1 4-fold axis of rotation	7	68	75–142	2	Tetragonal	P (simple)
							I (bc)[a]
Trigonal	1 3-fold axis of rotation	5	7	143–167	1	Rhombohedral (trigonal)	R (simple)
			18			Hexagonal	
Hexagonal	1 6-fold axis of rotation	7	27	168–194	1		P (simple)
Cubic	3 4-fold axes of rotation	5	36	195–230	3	Cubic	P (simple)
							I (bc)[a]
							F (fc)[c]

[a]*bc* body-centered, [b]base-centered, [c]*fc* face-centered. Note that seven crystal systems differ from six crystal families in that the hexagonal crystal family is split into two subsets: trigonal and hexagonal. The trigonal crystal system is larger than the rhombohedral lattice system, while the hexagonal crystal system is smaller than the hexagonal lattice system. The hexagonal crystal family is split into rhombohedral hexagonal lattice system.

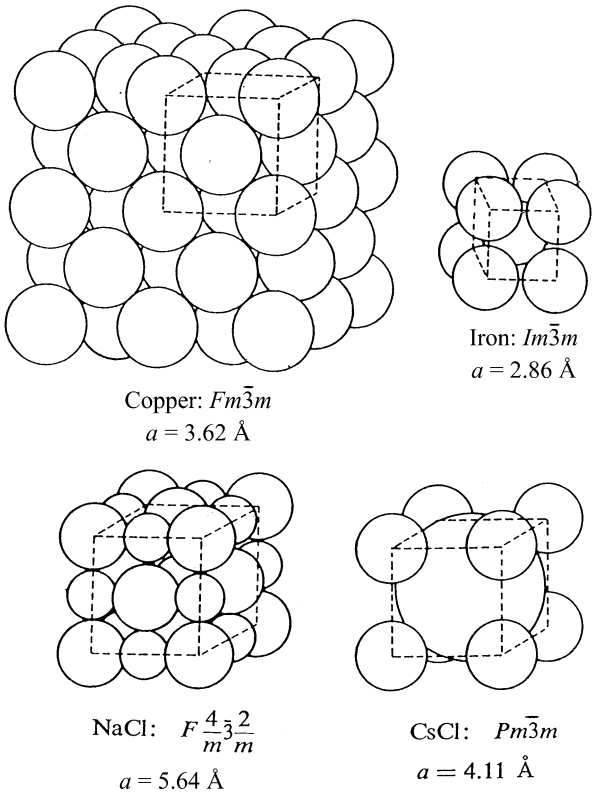


Fig. 3 Examples of cubic crystal unit cells and corresponding space group designations. The unit cell dimension (lattice parameter, *a*) is indicated. The edges of the unit cell are indicated by *dashed lines* which pass through the center of the corner atoms

crystal geometry implicit in the space group designations, Strukturbericht designations provide common composition information. However, the Strukturbericht designations often do not follow simple nomenclature evolution and can become confusing.

As noted in chapter ▶ Chemical Forces: Nanoparticles in the development of close-packed polyhedra, the same structure can be represented in a number of ways. This is particularly true for the fcc crystal structure unit cell as shown in ▶ Fig. 2 of chapter "Chemical Forces: Nanoparticles". Similarly, there are various possible ways to represent the unit cells for the fcc unit cell as shown in Fig. 5. In Fig. 5, these variations are not to be confused with the concept of polymorphism described above since only the regular fcc unit cell with four atoms per cell clearly exhibits the full symmetry of the arrangement ($Fm\bar{3}m$).

The closest approach distance between neighboring atoms in the fcc unit cell shown in Fig. 3 for Cu is $a\sqrt{2}$. Diffraction (X-rays or electrons) peaks or intensities are indexed such that the lattice spacing or cell dimension, a, is associated with the average spacing of corner atoms and larger atom spacing between neighboring atoms of the other possible cells represented in Fig. 5. Unlike the coordination of the close-packed polyhedral equivalents in ▶ Fig. 2 of chapter "Chemical Forces:

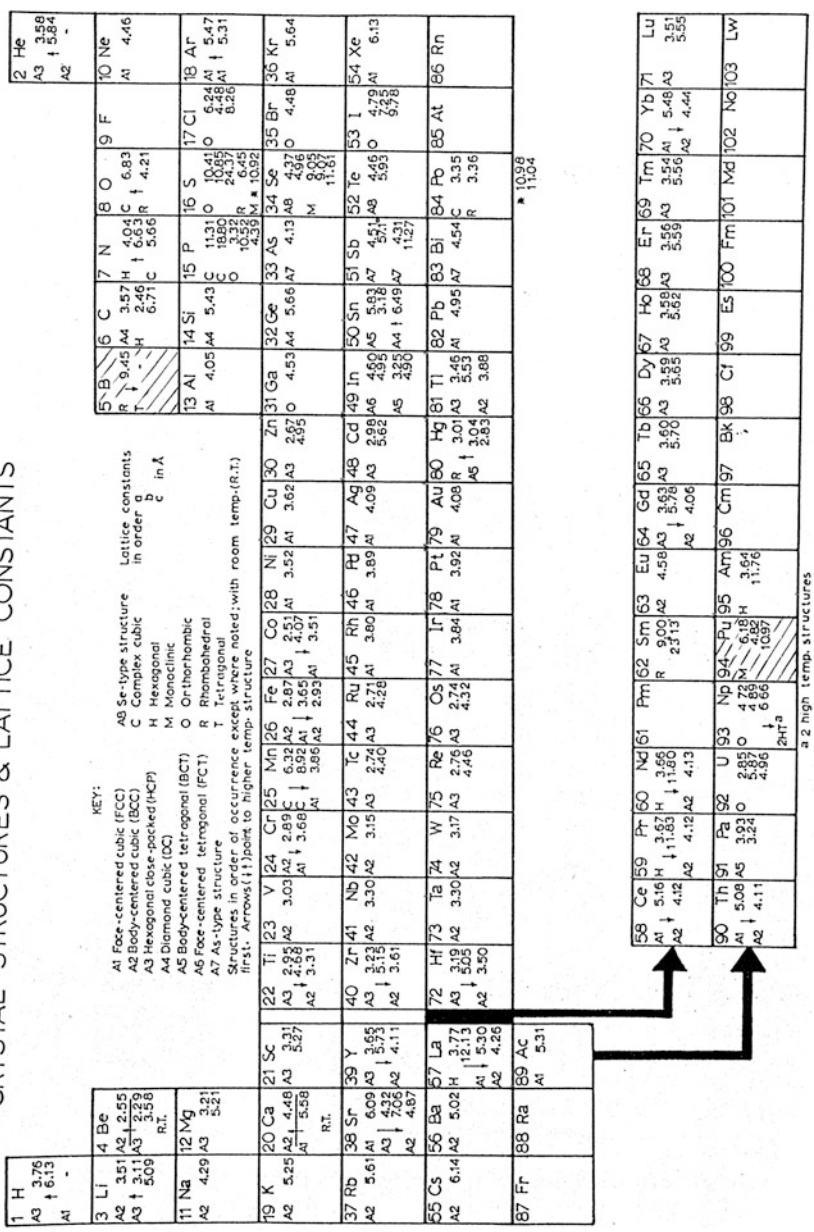


Fig. 4 Crystal structures and lattice parameters for the elements (From Staudhammer and Murr (1973))

Table 2 Crystal structure types and corresponding Strukturbericht symbol

Crystal structure type (composition)	Strukturbericht symbol
Elements in the periodic table (▶ Fig. 4 of chapter "Electromagnetic Color and Color in Materials")	A
AB compounds, AX compounds	B
AB one metal, X, O, F, S	
AB_2 compounds, AX_2 compounds	C
A_mB_n, A_mX_n compounds	D
ABX_m, $A_mB_nX_o$, or more complex compounds	E to K
Metal alloys	L
Organic compounds	O
Silicates	S

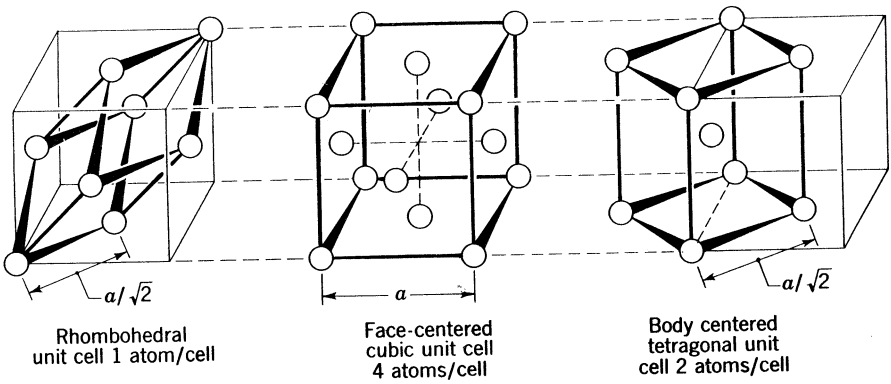


Fig. 5 Unit cell alternatives for the fcc structure (Adapted from Barrett and Massalski (1966))

Nanoparticles", only the fcc unit cell in Fig. 5 has a coordination number of 12 (12 equivalent neighbor distances). However, the hcp unit cell structure also exhibits a coordination number of 12. For bcc, the coordination number is 8, representing the equispaced corner atoms relative to the body-centered atom.

References

Barrett CS, Massalski TB (1966) Structure of metal, 3rd edn. McGraw-Hill, New York

Bravais A (1850) Memoire sur les systemes forms par despoints distributes regulierement sur un plan ou dans l'espace. J Ecole Polytech Paris 19:1

Buerger MJ (1960) Crystal structure analysis. Wiley, New York; See also Buerger MJ (1956) Elementary crystallography. Wiley, New York

Lonsdale K International table for X-ray crystallography (1952–1960) Vols 1–3, Kynoch Press, Birmingham. See also point and space groups symmetry data and complete space group tables on the WEB

Staudhammer KP, Murr LE (1973) Atlas of binary alloys. Marcel Dekker, New York

Lattice Directions and Planes, and Diffraction by Crystals

Contents

Abstract

Chapter "▶ Lattice Directions and Planes, and Diffraction by Crystals" reviews Miller index notation for crystal planes and directions as a precursor to establishing simple principles for diffraction by crystals. Bragg's law establishes diffraction geometry, while atomic scattering from specific atoms and structure factors determine the amplitudes of diffraction depending upon the atomic species and their crystal or lattice structure. The concept of the Ewald sphere and a reciprocal lattice as it relates to diffraction and diffraction amplitudes is also presented.

Introduction

The notations for lattice (or crystal) directions and planes (correspondingly atomic planes in crystals) are an important feature of crystallography as applied to orientations within crystals or crystal geometries. With reference to Fig. 1 which is somewhat specific to orthogonal lattices, ($\alpha = \beta = \gamma = 90°$) can be generally applied to unit cells in the context of ▶ Fig. 1c of chapter "Crystallography Principles." In Fig. 1, the position vector **r** directed from the origin at 000 through any lattice point within the unit cell (dotted) with position coordinates u_j, v_j, w_j is given by

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_10

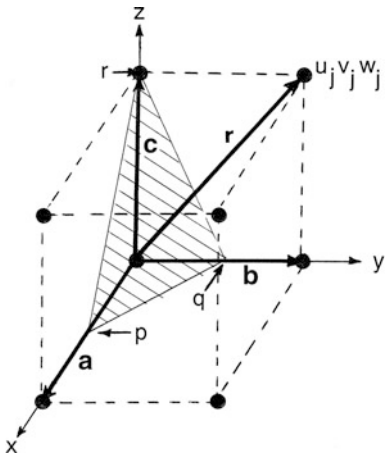


Fig. 1 Notations for directions and planes in a lattice unit cell. The jth atom at the unit cell corner is denoted ujvjwj. Shaded plane has intercepts along **a** (x), **b** (y), and **c** (z) denoted p, q, r

$$\mathbf{r} = u_j\boldsymbol{\alpha} + v_j\boldsymbol{b} + w_j\mathbf{c} \tag{1}$$

referred to as a real lattice vector, which characterizes a corresponding direction [uvw], which is the direction of a line from the origin to the lattice point, or a unit cell atom position in the crystal unit cell. Because unit cell dimensions are based on $|\mathbf{a}| = |\mathbf{b}| = |\mathbf{c}| = 1$ (unity), points or atoms located at fractions of the unit cell dimensions with fractional coordinates are reduced to the least common denominator, and the numerator assumes the designated direction indices: the face-centered atom in the unit cell base in Fig. 1 would have position coordinates $\left\lfloor \frac{1}{2}\frac{1}{2}0 \right\rfloor$ or [110]. Other directions of this form, [101] and [011], are designated directions of a form: <110>. That is, specific directions are denoted [uvw], while directions of this form are denoted <uvw>.

Planes in the lattice are determined by $(\mathrm{hkl}) = \mathrm{h} : \mathrm{k} : \mathrm{l} = \left(\frac{1}{\mathrm{p}}, \frac{1}{\mathrm{q}}, \frac{1}{\mathrm{r}}\right)$ (▶ Fig. 6 of chapter "Electromagnetic Color and Color in Materials"), and as in the case of directions, for planes outside the unit cell where p, q, and r can be >1, the indices are reduced to the lowest common denominator, and only the numerator expresses the indices of the plane, (hkl). This notation was devised by William Miller in 1839, and the indices h, k, l are commonly referred to as the Miller indices. In ▶ Fig. 6 of chapter "Electromagnetic Color and Color in Materials," if p = q = ½, r = 1, the plane is denoted (Eq. 1). Correspondingly planes of this form ((212), (122), and negative index planes) are denoted by planes of a form: {hkl}, since for cubic crystals implicit in Fig. 1, planes are perpendicular to directions and have equivalent indices. It is common practice to denote both planes and directions by (hkl): [hkl] and {hkl}: <hkl>.

In the case of directions, negative indices can characterize any coordinate in the unit cell equivalents in the four positive quadrants (+z) and four negative quadrants (−z). For planes and directions in cubic crystals, the indices for negative

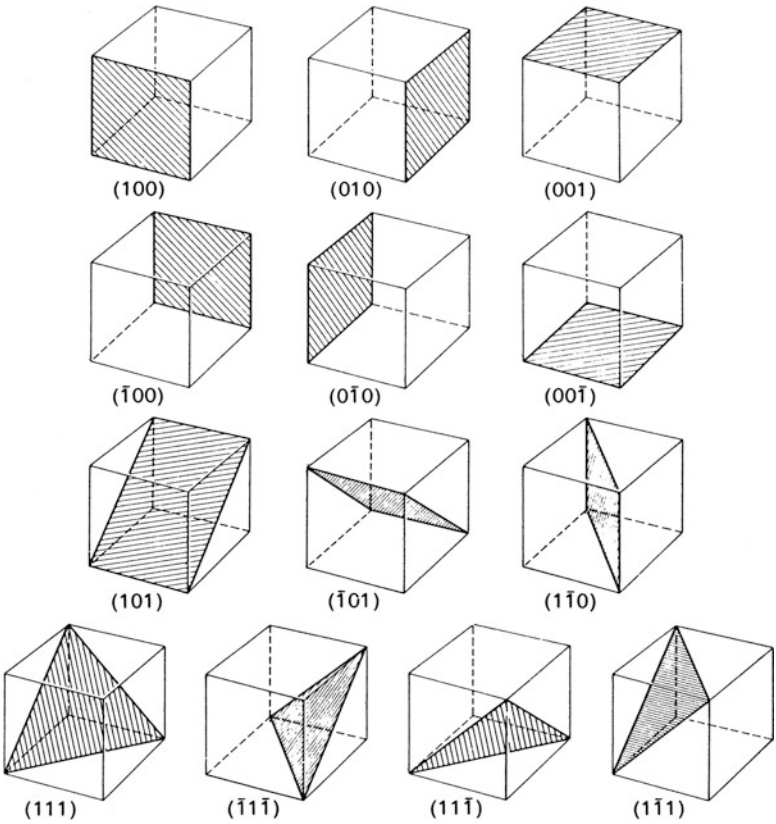


Fig. 2 Some common crystal planes in cubic crystals, including those corresponding to negative indices

coordinates or intercepts are designated with a bar above the index: [110] has a negative (or opposite direction) denoted $[\overline{1}10]$. Some examples of crystal plane designations, including negative indices for cubic unit cells, are illustrated in Fig. 2.

From the cube (orthogonal) geometry illustrated in the convention of Fig. 1, it can be observed that the distance between planes of the same form is given by

$$d_{hkl}(\text{orthogonal}) = \frac{1}{\sqrt{(h/a)^2 + (k/b)^2 + (l/c)^2}} \tag{2}$$

which for cubic crystals where $a = b = c$ becomes

$$d_{hkl}(\text{cubic}) = \frac{a}{\sqrt{h^2 + k^2 + l^2}} \tag{3}$$

Similarly, we can find that the angle between directions, θ, is given by

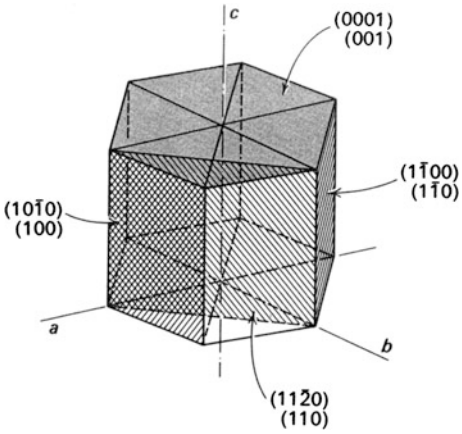


Fig. 3 Crystal planes in the hexagonal system

$$\theta = \cos^{-1}\left[\frac{h_1h_2 + k_1k_2 + l_1l_2}{\sqrt{h_1^2 + k_1^2 + l_1^2}\sqrt{h_2^2 + k_2^2 + l_2^2}}\right] \tag{4}$$

for $[h_1k_1l_1]$ and $[h_2k_2l_2]$ in a cubic crystal. Since in cubic crystals planes and directions are mutually perpendicular, the angle between two planes, $(h_1k_1l_1)$ and $(h_2k_2l_2)$, is given by Eq. 4. More general relationships for other crystal unit cells, including other orthogonal unit cells (▶ Fig. 2 of chapter "Crystallography Principles"), can be found in Murr (1991).

As shown in the periodic table of ▶ Fig. 4 of chapter "Crystallography Principles," many elemental metals have the hcp (A3) crystal structure. Indices based on the three axes, a, b, c in Fig. 1 using the Miller index notation are frequently used for hexagonal crystals, but all equivalent planes do not have similar indices, and the same can apply to directions. This can be avoided by using a four-index notation: (hkil), where $i = -(h + k)$. Figure 3 illustrates this index feature for hexagonal crystal unit cells, often called the Miller-Bravais notation.

Principles of Diffraction by Crystals

Prior to the development of diffraction principles by W. H. and W. L. Bragg around 1912, culminating in Bragg's law (▶ Eq. 4 of chapter "A Brief Introduction to Quantum Mechanics") for X-ray reflection from crystal planes of atoms, the geometry and dimensions of crystal unit cells were unknown. But Bragg's law only relates the angle of reflection (or diffraction) of X-rays and other radiation to the interplanar spacing of atoms (Eq. 2) providing dimensions or lattice parameters but no information regarding atom positions specific to the Bravais lattices (▶ Fig. 2 of chapter "Crystallography Principles") and crystal structures unique to specific space groups (▶ Table 1 of chapter "Crystallography Principles"). The fundamentals of the relationship between atom positions in the unit cells and diffraction intensities are

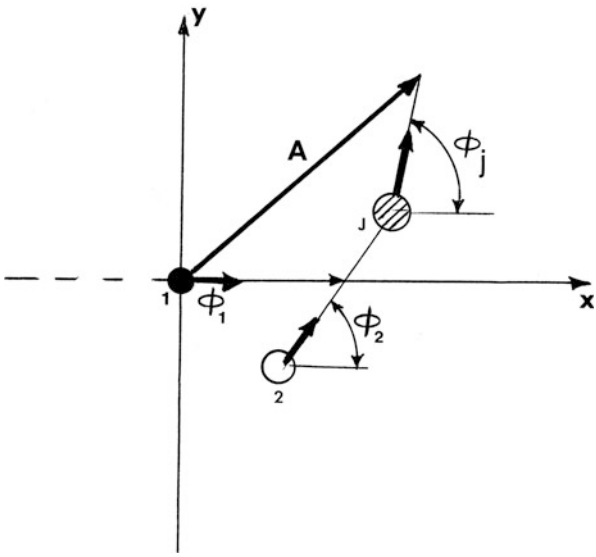


Fig. 4 Diffraction amplitude from a collection (or volume) of atoms (j) each scattering at a phase angle ϕ_j as represented in a simple 2D (x-y) geometry

implicit in the structure factor, a mathematical algorithm which relates real lattice points or atom positions in the crystal unit cell with intensity profiles at points representing a reciprocal lattice. This is a mathematical construct which relates the intensities and positions of the diffracted beams to atom positions and crystallography as well as the real sizes of crystalline matter, including the periodicity or range of periodicity of unit cells. In this context, we will present a brief survey of diffraction principles and fundamentals.

Radiation in the form of X-ray photons, electrons, or neutrons composing a beam or diagnostic probe will be selectively scattered by atoms in a diffracting volume of matter. This scattering depends upon the atom (specifically the atomic number, Z), the concentration of specific atoms, and their positions in the unit cell or cells composing the diffracting volume. Each atom scatters some amplitude of radiation determined by an atomic scattering factor, $f_j(\mu)$ for the jth atom represented in a random collection of different atoms shown schematically in Fig. 4. Each atom in Fig. 4 scatters some amplitude at a phase angle represented by ϕ_j for the jth atom. These scattering amplitudes produce diffracted beams having a resultant amplitude represented by

$$A = \sum\nolimits_j f_j(\mu)\exp\left(i\phi_j\right) \tag{5}$$

If we superimpose the Bragg diffraction geometry on Fig. 4, we can consider only two atoms characterizing two crystal planes separated by an interplanar spacing, d_{hkl}, which represents the scalar value of the real lattice vector, **r**, represented in Fig. 1 (and Eq. 1). This diffraction geometry is represented in Fig. 5. The incident beam of radiation is denoted by a unit vector $\mathbf{K}_o$, while the

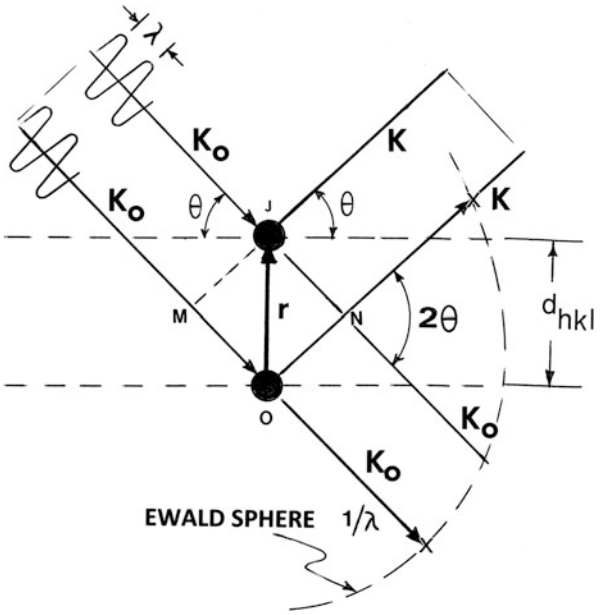


Fig. 5 Vector geometry for diffraction from a crystal unit cell (Fig. 1). Note the Ewald sphere boundary having a radius of 1/λ in reciprocal space

resultant, diffracted beam is denoted by a unit vector, **K**. Considering the diffraction geometry in Fig. 5, the path difference is given by

$$\Delta = \overline{OM} + \overline{ON} = \mathbf{r} \cdot (\mathbf{K} - \mathbf{K}_o) \tag{6}$$

and the corresponding phase difference is given by

$$\phi_j = \frac{2\pi}{\lambda}\Delta = \frac{2\pi}{\lambda}\mathbf{r} \cdot (\mathbf{K} - \mathbf{K}_o) \tag{7}$$

where it can be recognized from the scalar form of Bragg's law that

$$\Delta = 2d_{hkl}\sin\theta = \lambda \tag{8}$$

Consequently, the condition for diffraction is stipulated for $\Delta = \lambda$ or $\phi_j = 2\pi$. Examination of Eq. 7 with these conditions for diffraction results in

$$\frac{1}{\lambda}(\mathbf{K} - \mathbf{K}_o) = \frac{1}{\mathbf{r}} \tag{9}$$

which implies that the incident beam and diffracted beam have scalar values given by 1/λ which is self-consistent with the right-hand side of Eq. 9: 1/**r** whose scalar value is $1/d_{hke}$, implicit in Fig. 5. These reciprocal units necessitate a reciprocal lattice construct which can be fundamentally defined by

$$\mathbf{a}^* = 1/\mathbf{a};\ \ \mathbf{b}^* = 1/\mathbf{b};\ \ \mathbf{c}^* = 1/\mathrm{c} \tag{10}$$

or correspondingly

$$\mathbf{a}^* \cdot \mathbf{a} = 1 = \mathbf{b}^*\mathbf{b} = \mathbf{c}^*\mathbf{c}. \tag{11}$$

It follows that if the real crystal unit cell has a volume

$$\mathrm{V_c} = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = (\mathbf{b} \times \mathbf{c}) \cdot \mathbf{a} = (\mathbf{c} \times \mathbf{a}) \cdot \mathbf{b} \tag{12}$$

then the reciprocal lattice vectors can be expressed by

$$\mathbf{a}^* = (\mathbf{b} \times \mathbf{c})/\mathrm{V_c};\ \ \mathbf{b}^* = (\mathbf{a} \times \mathbf{c})/\mathrm{V_c};\ \ \mathbf{c}^* = (\mathbf{a} \times \mathbf{b})/\mathrm{V_c} \tag{13}$$

In the construction of the reciprocal lattice unit cell as shown in ▶ Fig. 11 of chapter "Electromagnetic Color and Color in Materials," the reciprocal lattice vector ($1/\mathbf{r}$) is expressed by

$$1/\mathbf{r} = \mathbf{g} = \mathrm{h}\mathbf{a}^* + \mathrm{k}\mathbf{b}^* + \mathrm{l}\mathbf{c}^* = \mathbf{g}_{\mathrm{hkl}} \tag{14}$$

where $\mathbf{g}$ or $\mathbf{g}_{\mathrm{hkl}}$ is referred to as a reciprocal lattice vector, and we can rewrite Eq. 9 as

$$(\mathbf{K} - \mathbf{K}_\mathrm{o}) = \lambda\mathbf{g} \text{ or } \mathbf{g} = \frac{1}{\lambda}(\mathbf{K} - \mathbf{K}_\mathrm{o}). \tag{15}$$

This illustrates that, as shown in Fig. 5, diffraction is satisfied when the Ewald sphere (with radius $1/\lambda$) passes through a reciprocal lattice point (shown by x in Fig. 5).

The Structure Factor

The amplitude of diffraction is now considered from a unit cell by considering Eq. 6 in the form

$$\mathrm{F(hkl)} = \mathrm{A} = \sum_\mathrm{j} \mathrm{f_j}(\mu)\mathrm{e}^{\mathrm{i}\phi_\mathrm{j}} \tag{16}$$

or from Eqs. 8 and 14

$$\mathrm{F(hkl)} = \sum_\mathrm{j} \mathrm{f_j}(\mu)\mathrm{e}^{2\pi\mathrm{i}\mathbf{r}\cdot\mathbf{g}}, \tag{17}$$

or

$$\mathrm{F(hkl)} = \mathrm{A} = \sum_\mathrm{j} \mathrm{f_j}(\mu)\mathrm{e}^{2\pi\mathrm{i}\left(\mathrm{hu_j}+\mathrm{kv_j}+\mathrm{lw_j}\right)} \tag{18}$$

where Eq. 18 is referred to as the crystal structure factor.

For X-rays, the atomic scattering factor is expressed generally by

$$F_{x(j)}(\mu) = \int_0^{\infty} 4\pi r^2 \rho(r) \frac{\sin \mu r}{\mu r} dr \tag{19}$$

where

$$\mu = \frac{4\pi \sin\theta}{\lambda} = 2\pi |\mathbf{g}| = 2\pi g \tag{20}$$

and $\rho(r)$ is related to the electron density of the atom (the atomic number, Z).

For electrons,

$$f_{e(j)}(\mu) = \frac{8\pi^2 me^2}{h^2} \left(\frac{z - f_{x(j)}(\mu)}{\mu^2} \right) \tag{21}$$

or

$$f_{e(j)}(\mu) = \frac{m}{2} \left(\frac{e}{h} \right)^2 \left(\frac{z - f_{x(j)}(\mu)}{(\sin\theta/\lambda)^2} \right). \tag{22}$$

Values of the atomic scattering factor for X-rays ($f_{x(j)}(\mu)$) and electrons ($f_{e(j)}(\mu)$) have been tabulated as a function of $\sin\theta/\lambda$ for atomic number, Z (Barrett and Massalski 1966; Murr 1991; Vainshtein 1964).

The structure factor in Eq. 18 can be interrogated for any crystal structure or crystal unit cell in order to determine amplitudes of diffraction for each diffracting plane (hkl). It can be observed in Fig. 7 that diffraction from crystal planes {hkl} in the real crystal is represented by diffraction vectors ideally perpendicular to these planes by points in the reciprocal lattice, which have position coordinates h, k, l in contrast to position coordinates of atoms in the real crystal lattice: u_j, v_j, w_j. This is a consequence of the cross products of real lattice vectors as defined for the reciprocal lattice in Eq. 13, with reference to Fig. 6. Correspondingly, crystal planes, {hkl}, in the real lattice become points in the reciprocal lattice or vectors perpendicular to these planes, especially for cubic crystals.

It can be easily demonstrated that there are no diffraction extinctions for simple cubic crystals (or unit cells), and all planes will diffract since $F(hkl) = f_j(\mu)$. For bcc crystal unit cells, the body-centered atom has coordinates $\frac{1}{2}\frac{1}{2}\frac{1}{2}$, and the structure factor becomes

$$F(hkl)_{bcc} = f_j(\mu) \left[1 + e^{\pi i(h+k+l)} \right] \tag{23}$$

(using 000 for $u_j\, v_j\, w_j$ for corner atom symmetry and 8 × 1/8 atoms per corner, etc.)

Therefore, it is observed that if (h + k + l) is an odd integer, the structure factor will be zero, and the intensity of diffraction from corresponding crystal planes will be zero:

$$|F(hkl)|^2 = 0 \text{ for bcc} \tag{24}$$

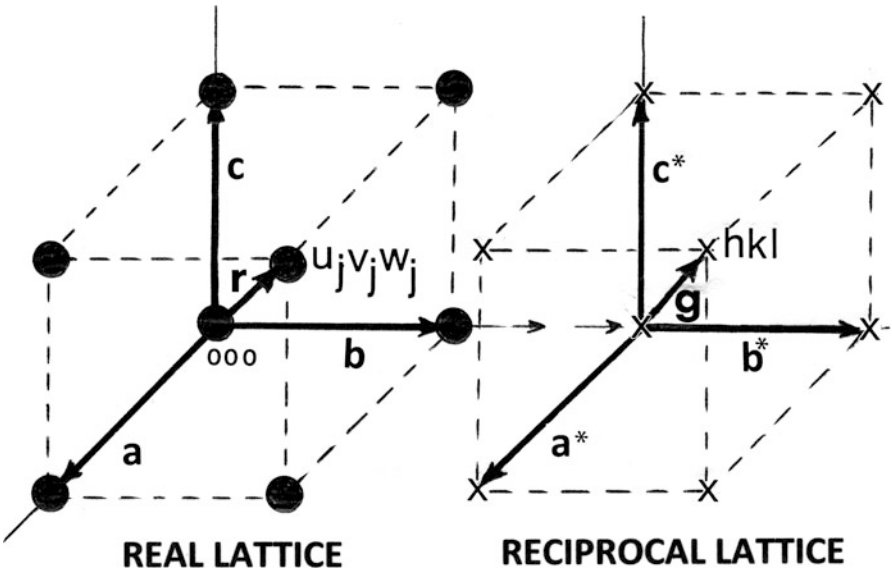


Fig. 6 Comparison of the real (crystal) lattice unit cell with the reciprocal lattice

when $(h + k + l) = 1, 3, 5\ldots$.

For fcc crystal unit cells, the structure factor becomes (for corner and face atoms: $\frac{1}{2}\frac{1}{2}0$, etc.)

$$F(hkl)_{fcc} = f_j(\mu)\left[1 + e^{\pi i(h+l)} + e^{\pi i(k+l)} + e^{\pi i(h+k)}\right]. \tag{25}$$

Consequently, for diffraction to be extinct ($F(hkl)_{fcc} = 0$), $(h + l)$, $(k + l)$, and $(h + k)$ present mixed even and odd integer conditions.

The diffraction conditions implicit in Eqs. 23 and 25 for bcc and fcc crystal unit cells apply to single atom species. For ionic (bcc) crystal unit cells such as the CsCl unit cell in ▶ Fig. 3 of chapter "Crystallography Principles," the structure factor becomes

$$F(hkl)_{CsCl} = f_{Cs}(\mu) + f_{Cl}(\mu)e^{\pi i(h+k+l)} \tag{26}$$

which produces different diffraction intensities for specific, diffracting planes {hkl}.

Table 1 summarizes the order of diffraction planes for bcc, fcc, diamond cubic, and hcp crystal structures of the same atomic species. These correspond to increasing values of the reciprocal lattice vectors which for cubic crystals would be

$$g = \sqrt{h^2 + k^2 + l^2}/a = 1/d_{hkl} \tag{27}$$

Table 1 Allowed diffraction planes {hkl} for single-species crystal unit cells shown

bcc {hkl}	fcc {hkl}	Diamond {hkl}	hcp {hkl}
110	111	111	100
200	200	220	002
211	220	311	101
220	311	400	102
310	222	331	110
322	400	422	103
321	331	511, 333	200
400	420	440	112
411, 330	422	531	201
420	511, 333	620	004
332	440	533	202
422	531	444	104
510, 431	600, 442	711, 551	203
521	620	642	210
440	533	553	211

Ideally, this sequence of diffracting planes would correspond to diffraction peaks in X-ray diffractometer spectra or the sequence of diffraction rings in electron diffraction patterns for small crystal or particle sizes. These features are illustrated in Fig. 8 for nickel which is fcc with a lattice parameter, a = 3.52 Å. Note the correspondence between the X-ray diffraction (XRD) peaks in Fig. 7a (marked 1 to 5) and the electron diffraction rings in Fig. 8b (marked 1 to 5). In Fig. 7a, the diffracting planes and lattice parameter correspond to the interplanar spacing (in Eq. 3) where

$$d_{hkl} = \lambda/2 \sin \theta \tag{28}$$

and λ is the X-ray wavelength for X-rays produced from a copper target (Cu-Kα X-rays) and equal to 1.54 Å. In Fig. 7b, the diffracting planes and lattice parameter are related to the same interplanar spacing:

$$d_{hkl} = \lambda L/R, \tag{29}$$

where λ is the electron wavelength in the electron microscope operating at 300 kV and equal to 0.2 Å. L in Eq. 29 is the electron microscope camera length (in centimeters), and λL is commonly referred to as the *camera constant*, determined by calibrating the instrument with a known standard (such as a nickel thin film). R in Eq. 29 is the radius of the diffraction rings in Fig. 7b measured in centimeters.

In contrast to diffraction from a powder mixture having random orientations as shown by XRD in Fig. 7a or a polycrystalline thin film as in Fig. 7b for electron diffraction (referred to as selective-area electron diffraction or SAED), diffraction

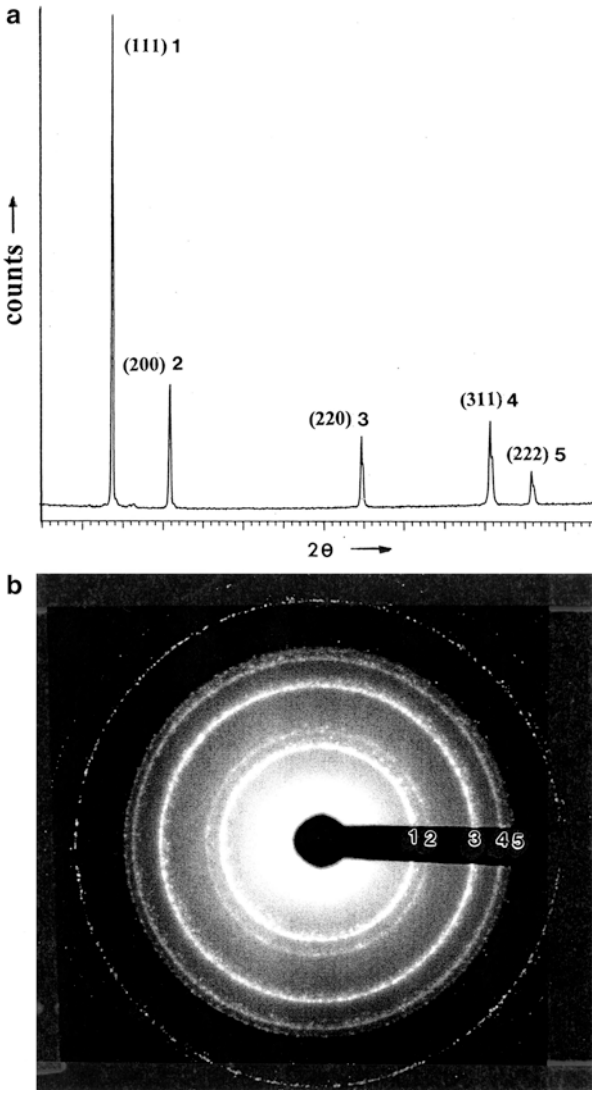


Fig. 7 Diffraction from random particles and polycrystals of nickel. (**a**) XRD spectrum using Cu-Kα x-radiation. (**b**) Electron diffraction pattern from a polycrystalline thin foil in the TEM operating at 300 kV. Diffracting planes are noted 1 to 5 in both (**a**) and (**b**)

from a single crystal or a set of oriented grains all having the same diffracting plane (hkl) will exhibit a single spectral peak in XRD or a specific arrangement of diffraction spots in the SAED pattern. Figure 8 shows diffraction from an fcc NiCr alloy (a = 3.56 Å) having a [100] grain surface orientation. Since diffraction from {100} is forbidden, the XRD peak in Fig. 8a shows the (200) diffracting plane, parallel to (100), and the half-cell plane. Correspondingly, the SAED pattern in Fig. 8b shows a single-crystal-oriented [100], with diffraction spots from allowed diffracting planes marked (see Table 1). The diffraction spots in Fig. 8b represent

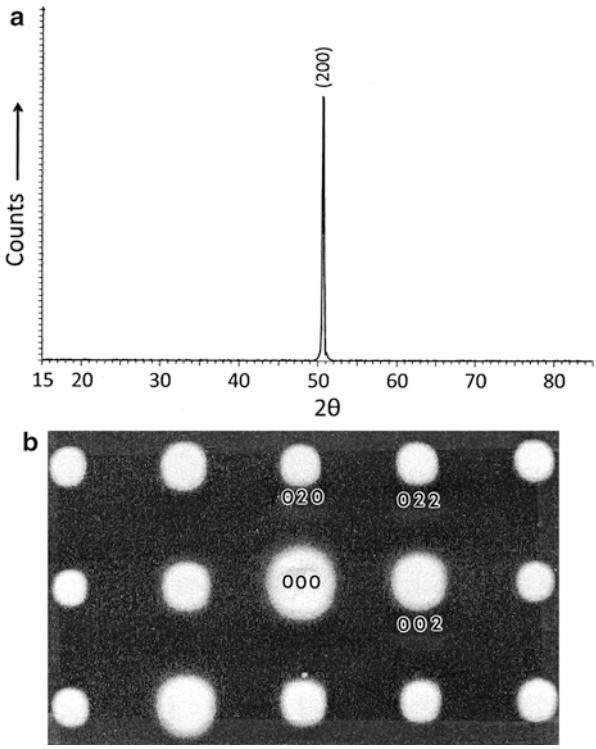


Fig. 8 Diffraction from [100] single crystals of fcc NiCr alloy. (**a**) XRD spectrum (Cu-Kα x-radiation). (**b**) [100] fcc electron diffraction pattern in TEM at 300 kV accelerating potential

reciprocal lattice vectors referenced to the origin at 000. These are determined from the zone axis criterion

$$hH + kK + lL = 0 \tag{30}$$

where [HKL] is the zone axis, and [hkl] are the reciprocal lattice vectors normal to this axis in the electron microscope. Consequently, for [HKL] = [100], the condition for diffraction is h = 0 or (0kl). Therefore, from the listing of the sequence of allowed diffracting planes for fcc in Table 1, we can have 020, 022, and 002 (and the negatives for opposite diffraction spots: $0\bar{2}0$, $0\bar{2}\bar{2}$, and $00\bar{2}$). These are regarded as the first-order diffraction spots.

At this point, it should be recognized that while the intent of this book has focused on more advanced materials issues, we have, as illustrated above, injected fundamentals somewhat strategically. For the reader who may wish to take a more in-depth look at crystallography principles and expand on structure factors or diffraction principles, the International Tables for X-ray Crystallography (1952–1960) continue to be extensive, while the work of Barrett and Massalski (1966) and Murr (1991) can be supplemental and informative. Of course XRD and SAED techniques will be utilized in the continuing development of materials structures and properties that follow.

References

Barrett CS, Massalski TB (1966) Structure of metal, 3rd edn. McGraw-Hill, New York
International Table for X-ray crystallography, (1952–1960), vols 1–3, Kynoch Press, Birmingham. See also point and space grade symmetry data and complete space group tables on the WEB
Murr LE (1991) Electron and ion microscopy and microanalysis: principles and applications, 2nd edn. Marcel Dekker, New York
Vainshtein BK (1964) Structural analysis by electron diffraction. Pergamon Press, New York

Structure of Metals and Alloys

Contents

Abstract

There are roughly 70 metallic and metallike elements with correspondingly 2,485 binary combinations or alloys. This chapter begins with a brief introduction followed by an overview of boron (Z = 5), a metalloid exhibiting unusual polymorphic structures, and a comparative overview of plutonium, another unusual polymorphic metal near the end of the periodic table (Z = 94). These two elements represent perhaps the most unusual ranges of polymorphic structures and properties in the periodic table.

Introduction

Common, elemental metals have primarily cubic (fcc or bcc) or hexagonal (hcp) crystal structures (see ▶ Fig. 3a, b and the periodic table of Fig. 4 of chapter "Crystallography Principles"). Some of the metals are polymorphic such as Li, Na, Ti, Mn, Te, Co, etc., and as noted previously Pu is the most polymorphic metal having six different crystal structures, including fcc and bcc structures.

Metals, and metals in combinations forming alloys, intermetallics, or complex phases, do not normally require charge balancing or other compensation in the crystal lattice for substitution in atomic sites or as interstitials in the crystal lattice. There is only the preference, in some structures for energy minimization such as ordering of disordered atomic structures which can form *superlattice* arrangements in binary alloys such as CuAu or Cu–Zn. Examples of these structures are

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_11

illustrated in Fig. 1 for corresponding structure types shown in ▶ Table 2 of chapter "Crystallography Principles." Also represented in Fig. 1c is the change in resistivity from the *disordered* to *ordered state* which can occur upon cooling below a critical temperature (Tc) or narrow temperature range. The electron mean free path (▶ Eq. 1 in Chapter "Electromagnetic Color and Color in Materials") is increased in the ordered phase, and overall conductivity would be increased (with a corresponding decrease in resistivity) as the superlattice period is extended as shown in Fig. 1d. Boundaries are formed between superperiods and referred to as an antiphase boundary (APB) separating superlattice domains which become 2D interfaces in 3D superlattice structures.

Binary metal alloys such as noble metal alloys depicted in Fig. 1b, d maintain metallic bonding over a continuous range of solid solubility in their binary phase diagrams.

Other systems such as Ni_3Al or Ti_3Al represent *eutectic* (intermetallic) phases having some degree of covalent bonding and correspondingly unique or different properties: mechanical and electrical in particular. Such alloy systems are often referred to as intermediate alloys, and this mixed mode of bonding can occur between all types of bonding: ionic, covalent, metallic, and even weak bonding characteristic of van der Waals bonding. For example, binary metal alloys are also found to occur in NaCl- and CaF_2-type (Bl or Cl) ionic structures: MgSe, MnSe, and PbTe and $PtSn_2$, $SeCu_2$, Cu(BiMg), Cu(CdSb), and Ag(AsMg), to illustrate only a few examples.

As will be seen later, the NaCl and CaF_2 structures are also very common among oxides, sulfides, chlorides, fluorides, and carbides. Many of these, particularly oxides, are typically ionic sulfide crystal structures. Carbides often represent the intermetallics and are dominated by covalent crystal structures and many have tetragonal unit cells.

Of course there are a wide range of nonstoichiometric alloys having very complex, disordered crystal structures, and many such alloy systems are polymorphic. A typical binary system exhibiting structural polymorphs is Cu–Zn, illustrated in the binary phase diagram of Fig. 2a. Compositions of Cu–Zn can be bcc disordered or ordered, and disordered bcc structures can occur up to about 65Cu–35Zn, while compositions of 70 Cu–30 Zn are disordered fcc, and referred to as cartridge brass, or α-brass in contrast to bcc β-brass (or yellow brass). As observed from the phase diagram in Fig. 2a, mixtures of α and β phase can occur, and these brass phase mixtures are often referred to as Muntz metal or α-β brass. The shaded α-β region in the phase diagram of Fig. 2a shows this region; correspondingly, Fig. 3b shows a ternary isotherm for the Fe–Cr–Ni system having a large γ (fcc) phase region (shaded). Figure 3 illustrates the typical metallographic images for these brass structures in polycrystalline, commercial products. Of course these crystal phases and mixtures are easily discernible by XRD illustrated in ▶ Fig. 7a of chapter "Lattice Directions and Planes, and Diffraction by Crystals."

The ternary phase diagram in Fig. 2b characterizes popular stainless steels (Fc–18Cr–8Ni, in weight percent), as well as Ni–Cr-based alloys such as the fcc (γ) matrix in Inconels or related nickel-base superalloys. Such alloys resemble

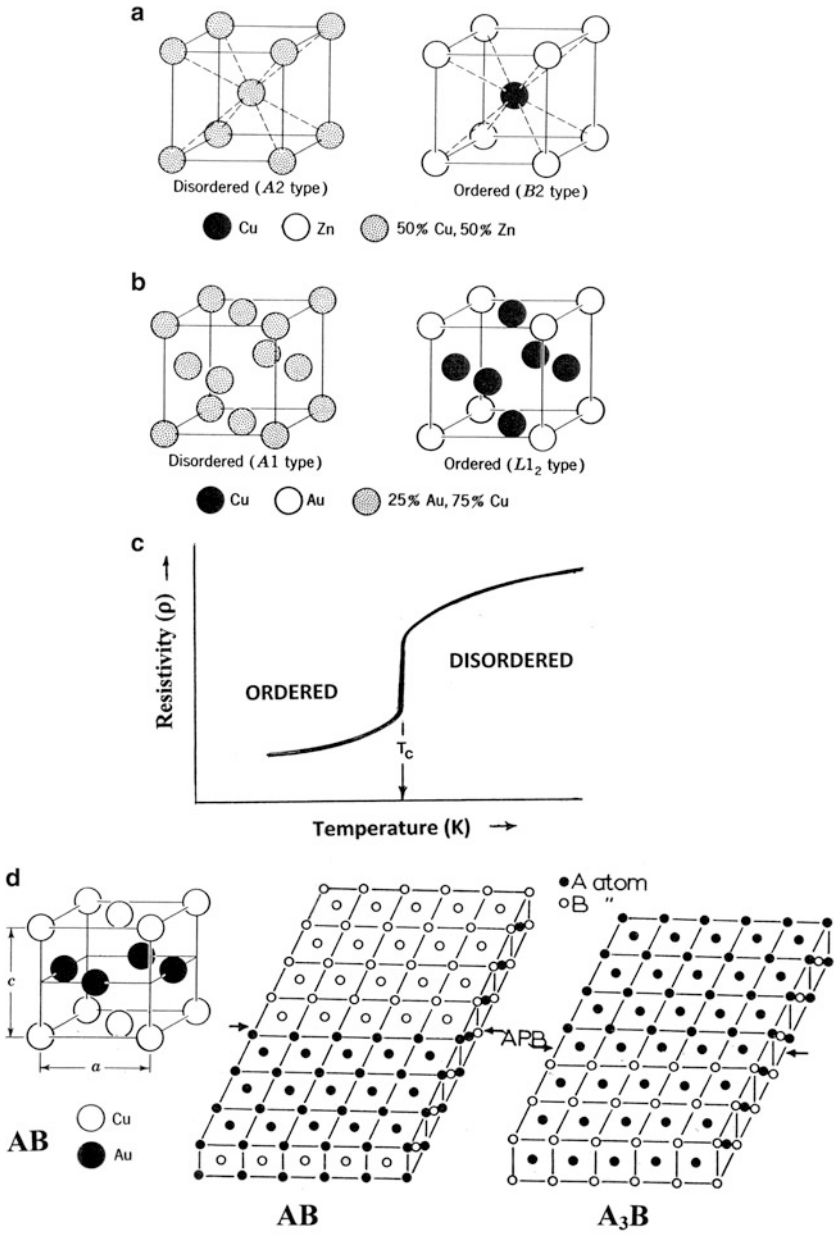


Fig. 1 Order–disorder structures creating superlattices (**a**) Cu–Zn (AB structure). (**b**) Cu–Au (A_3B structure) (Adapted from Barrett and Massalski (1966)). (**c**) Resistivity transformation for A_3B order–disorder. Tc = 663 K for Cu_3Au. (**d**) Long-period superlattice structures with an antiphase boundary (APB) separating AB and A_3B structure content A or B in (001). The AB structure at *right* is orthorhombic

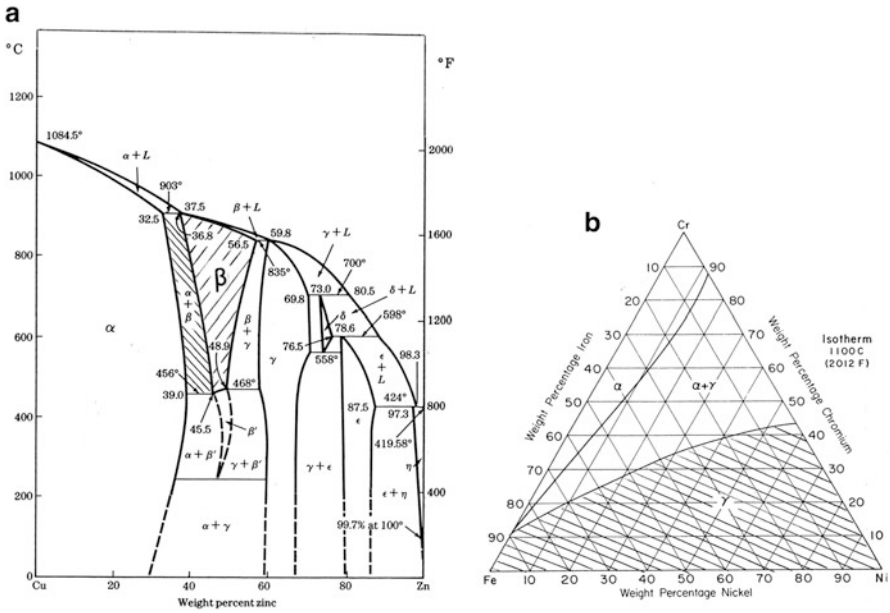


Fig. 2 Examples of popular binary and tertiary alloy phase diagrams. (**a**) Cu–Zn (brass compositions). (**b**) Fe–Cr–Ni (stainless steel compositions)

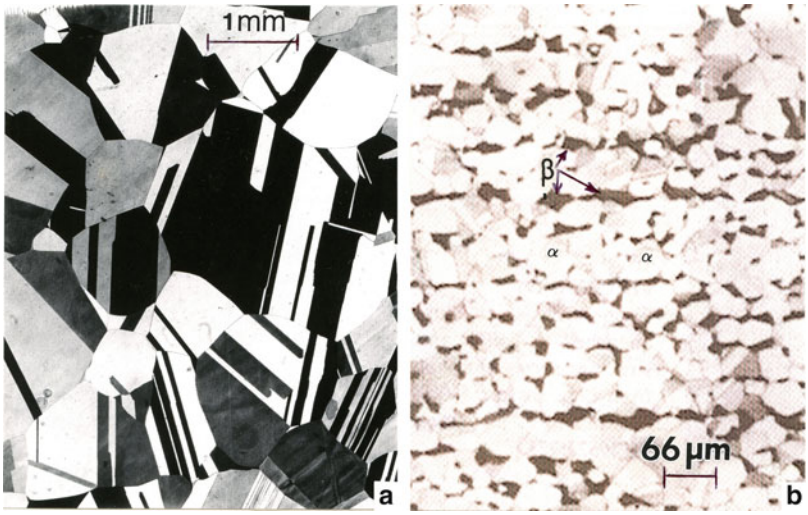


Fig. 3 Light microscope images for polycrystalline brass structures. (**a**) α-brass showing grains and annealing twins (straight and parallel boundaries). (**b**) α-β brass mixture (Muntz metal)

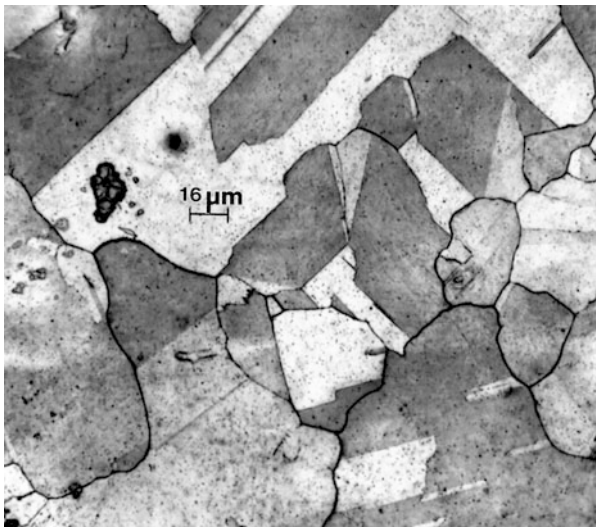


Fig. 4 Inconel 625 (fcc) grain structure showing equiaxed grains with annealing twin fcc boundaries (straight boundaries coincident with {fcc} planes. The fcc crystal unit cell lattice parameter is 3.59 Å. The composition is 64Ni–s1Cr–9Mo–4Nb (in weight percent))

α-brass since they contain a high fraction of annealing twin boundaries as shown typically in Fig. 4. Hume–Rothery (1936) rules for complete solid solubility include same crystal structure, <15 % difference in atomic radii, equivalent valence, and similar electronegativity. There is a tendency for all metals to form normal valence compounds with groups IVB (Ti), VB (V), and VIB (Cr) in the middle of the periodic table (▶ Fig. 4 of chapter "Crystallography Principles"). Examples include equilibrium diagrams for the noble metals: Cu, Ag, and Au with metals of the B subgroups. As noted above, Cu–Zn (Fig. 2a) is an example of these systems. Some structures are also related to the ratio of electrons to atoms present in the structure (e/a) and possess wide ranges of solid solubility. These are often referred to as *electron phases*. For example, for β-Cu–Zn (Fig. 2a), e/a = 3/2. These structural features are discussed in more detail in the classic work of Barrett and Massalski (1966).

There are a total of 5,356 possible binary (AX) combinations for the first 103 elements (Z = 1–103) in the periodic table of ▶ Fig. 4 of chapter "Crystallography Principles," including the inert or rare gases and other rare (or scarce) elements. More practical combinations for 86 more tractable elements yield 3,741 (86 (86 + 1)/2) combinations. There are roughly 70 readily available metallic elements with correspondingly 2,485 combinations of metals, AB, or binary alloys. Consequently, there will be more orders of magnitude of $A_m B_n$ (m, n = rational integers) combinations and many more orders of magnitude of other binary compounds and ternary and higher-order rational and irrational phases. In addition, and as illustrated generally in ▶ Fig. 4 of chapter "Crystallography Principles," many elements are polymorphic (or allotropic), and many binary and higher-order phases are polymorphic. It is estimated that currently (2013) more than 22 million chemical substances have been identified (see http://www.CAS.org/).

Consequently, it is only possible to provide a few appropriate, unusual, or important structural examples in this compendium.

To further exemplify these features, we will examine boron (Z = 5) and plutonium (Z = 94) shown shaded in ▶ Fig. 4 of chapter "Crystallography Principles." While not shown specifically in the periodic table of ▶ Fig. 4 of chapter "Crystallography Principles," boron has five crystal phases and an amorphous phase: α-rhombohedral (trigonal), β-rhombohedral, γ-orthorhombic, α-tetragonal, and β-tetragonal. Correspondingly, plutonium exhibits six crystal phases: α-Pu (simple monoclinic), β-Pu (body-centered monoclinic), γ-Pu (face-centered orthorhombic), δ-Pu (face-centered cubic), δ′-Pu (body-centered tetragonal), and ϵ-Pu (body-centered cubic). These elemental polymorphs along with a few examples of their common alloys and compounds provide a fascinating contrast of unusual material structures and properties representing disparate portions of the periodic table.

Boron Polymorphs and Compounds

Boron (Z = 5 shown shaded in the periodic table in ▶ Fig. 4 of chapter "Crystallography Principles") exhibits some of the most remarkable structures and properties of any element. First isolated in 1808 it is extremely unreactive at room temperature, extremely hard (~50 GPa microindentation hardness in contrast to ~5 GPa for steels and superalloys), lighter than aluminum, and extremely strong (Young's modulus ~450 GPa in contrast to ~150–210 GPa for steels and Co-base superalloys). Boron's melting point is 2,450 K although it remains liquid only within a narrow temperature range. Prominent material applications have included abrasives and cutting tools, thermoelectricity, and as a silicon dopant to create p-type semiconductors. Because it has a high neutron capture cross section, it has, along with key compounds, played a significant role as a thermal neutron absorber in control rods and shields for nuclear (fission) reactors. Although not a true metal, boron is often considered a metalloid.

All known crystal structures or polymorphs of boron contain icosahedral clusters similar to the 13-atom icosahedrons (▶ Fig. 2a of chapter "Chemical Forces: Nanoparticles"), but with no central atom. Consequently, this cage-like cluster, with a boron atom at each of the icosahedron vertices (often indicated as B_{12}), is often considered to be characterized as an icosahedral molecule. The B_{12} icosahedral cluster has three, metallic-like center bonds within the icosahedrons and two and three center bonds between icosahedral clusters in crystal structures. The ideal fivefold symmetry of the icosahedrons cannot propagate or fill space in the boron crystal lattice. In the icosahedral cluster (B_{12}) each boron forms a covalent bond with each of its five neighbors which occupy equivalent positions about a fivefold axis as noted. In addition to these 6 fivefold axes, there are ten threefold axes, 15 twofold axes, 15 mirror planes, and a center of inversion. Some refer to this icosahedral boron structure, shown in Fig. 5a, as a metalloid (Hoard and Hughes 1967).

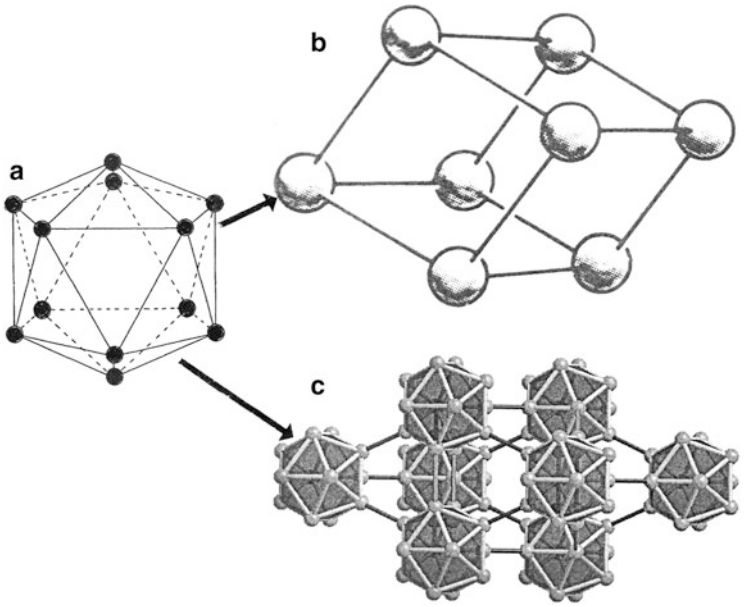


Fig. 5 α-rhombohedral (α-B_{12}) boron polymorph. (**a**) B_{12} icosahedron cluster. (**b**) Basis rhombohedral (trigonal) unit cell. (**c**) α-B_{12} rhombohedral structure (Adapted from Widom and Mihalkovic (2008))

Figure 5b, c illustrates the α-rhombohedral allotrope of boron which is stable up to 1,400 °C. The more stable allotrope, β-rhombohedral (trigonal) (up to 2,450 °C), is considerably more complex than the α-phase, and contains up to 320 B atoms in the full unit cell, but is normally referred to as β-B_{106}. Figure 6 provides some views of this structure, while Fig. 6 shows the γ-orthorhombic polymorphic cell structure containing 28 B atoms and often referred to as γ-B_{28}. Two other tetragonal allotropes exist: α-tetragonal and β-tetragonal sometimes referred to as T-50 and T-192 indicating 50 and 192 B atoms per unit cell, respectively. Table 1 summarizes the corresponding crystal structure data, hardness, and elastic modulus data for these boron polymorphs.

It can be noted in the projection views of the β-rhombohedral structure in Fig. 6 that it is a packing of B_{12} $(B_5)_{12}$ groups, although there are controversies regarding this structure, or structure variations, which have included B_{106}, B_{314}, and B_{320}. Figure 6 shows essentially the B_{106} primitive cell where the number of nearest neighbor atoms for each boron ranges from 4 to 8, which is very different from α-B_{12} (Fig. 5c), the orthorhombic (γ-B_{28}) structure shown in Fig. 7a, and the α-tetragonal (T-50) projection shown in Fig. 7b. The γ-B_{28} structure in Fig. 7a consists of icosahedral B_{12} clusters and B_2 pairs in an NaCl-type arrangement, with these two boron species occupying equivalent anionic and cationic sites. In this respect, the structure, unlike the α-B_{12} and β-B_{106}, is slightly ionic. In the α-tetragonal (T-50 or B_{50}) polymorph shown in the projection of Fig. 7b, the unit

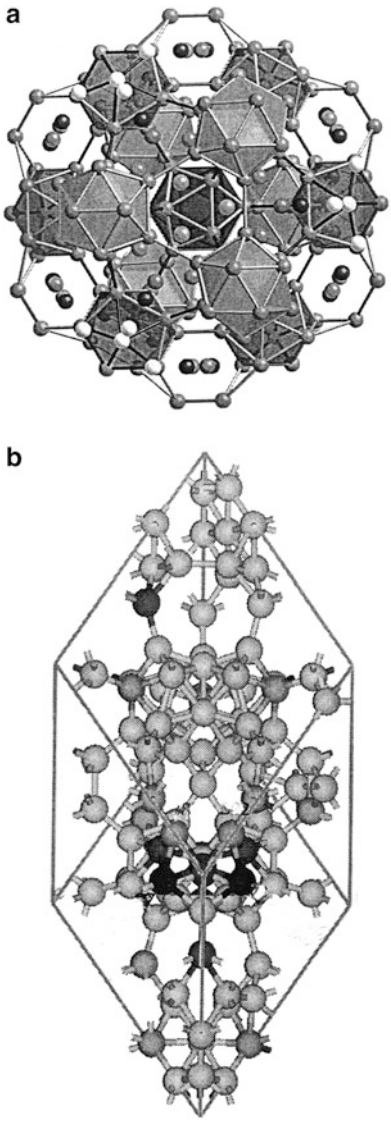


Fig. 6 Structure of β-rhombohedral boron. (**a**) View along rhombohedral (trigonal) [111] axis. (**b**) Primitive cell is adapted from Widom and Mihalkovic (2008)

cell contains four icosahedra and two isolated boron atoms: $(B_{12})_4$ $(B)_2$. The β-tetragonal (T-192 or B_{192}) phase shown in Table 1 consists of chains of B_{12} icosahedra alternately aligned in the a and b axes directions along with so-called twinned (or joined) double icosahedra (B_{21}) linked to 10 adjacent B_{12} icosahedra in these chains and to 4 neighboring double icosahedra, forming a rather complex icosahedral cluster structure.

It can be noted in Table 1 that the γ-B_{28} (orthorhombic) polymorph has the highest measured hardnesses and corresponding Young's modulus, which is

Table 1 Crystal structure data for boron

Phase	Unit cell parameters	Space lattice and space groups	Atoms per unit cell	Density (g/cm^3)	Hardness HV (GPa)	Modulus of elasticity E (GPa)
α (alpha) rhombohedral	$a = 5.06$ Å $\alpha = 58.06°$ (B-12)	Trigonal R $\bar{3}$ m (#166)	12	2.46	42	400
β (beta) rhombohedral	$a = 10.14$ Å $\alpha = 65.3°$	Trigonal R $\bar{3}$ m (#166)	(106)[a] (320)	2.35	45	470
δ (gamma) orthorhombic	$a = 5.05$ Å $b = 5.62$ Å $c = 6.97$ Å	Orthorhombic Pnnm (#58)	28	2.54	50–58	490
α (alpha) tetragonal	$a = 8.75$ Å $c = 5.06$ Å (T-50)	Tetragonal $P4_2/nnm$ (#134)	50	2.36	–	–
β (beta) tetragonal	$a = 10.12$ Å $c = 14.14$ Å (T-192)	Tetragonal P4, (#76)	192	2.36	–	–

Hardness and modulus of elasticity (Young's modulus) data is averaged from numerous sources
[a]Sub-cell. Full unit cell contains 320 atoms

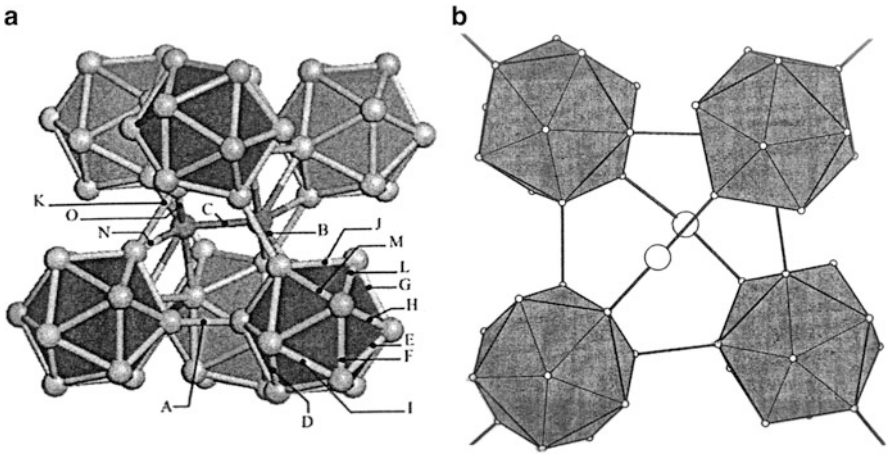


Fig. 7 (**a**) Orthorhombic $\gamma - B$ structure. (**b**) Crystal structure projection for α-tetragonal B_{50}. (**a**) is adapted from Klements et al. (1960). The 15 bonds within the B_{12} cluster, between clusters, and between the B_2 component are lettered A to O

comparatively lower hardness than diamond, which ranges from 70 to 150 GPa. However, boron nitride (cubic) (BN) with the zinc blende (or diamond cubic modification) structure shown in Fig. 8a can have hardnesses ranging from 48 to 62 GPa, and Young's modulus as high as 600 GPa, and exceeded only by

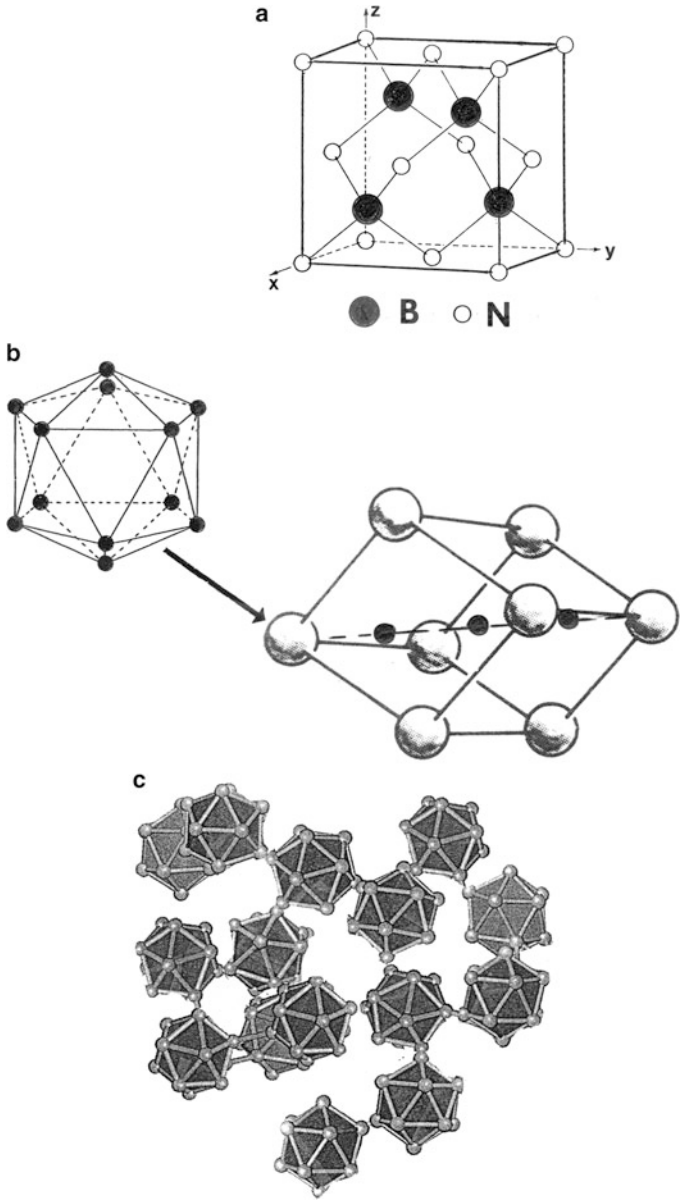


Fig. 8 Cubic (zinc blend (ZnS)) BH crystal structure (**a**). Rhombohedral (trigonal) B_4C ($B_{12}C_3$) crystal unit cell (**b**). Amorphous boron (simple schematic) (**c**)

ReB_2 (E = 700 GPa). However, in contrast to cubic BN in Fig. 8a B_4C has a rhombohedral, icosahedral (B_{12}) structure as shown in Fig. 8b and a hardness of only 30 GPa. The B_4C structure is actually $B_{12}C_3$ in Fig. 8b, corresponding to a chain of three boron atoms in the rhombohedral [111] axis. It may be recalled that

BN has a hexagonal polymorph shown in ▶ Fig. 6 of chapter "Chemical Forces: Molecules."

Boron and boron compounds also form whiskers and nanotubes which can have exceptionally high hardness and strength (elastic on Young's modulus (E)). As a consequence, these fibers and nanotubes find novel applications in the production of strong composites, especially replacing glass fibers in polymeric matrices as well as metal-matrix composites. Boron and many boron compounds also become amorphous, and boron as an amorphous phase assumes an interesting "glass" structure composed of irregular arrays of B_{12} icosahedra in contrast to the tetrahedral SiO_4 arrays composing normal glass structures illustrated in ▶ Fig. 3 of chapter "Electromagnetic Color and Color in Materials." Figure 8c illustrates this amorphous boron (glass) concept. In some respects, the tetrahedral arrays in ▶ Fig. 3 of chapter "Electromagnetic Color and Color in Materials" and the icosahedral arrays in Fig. 8c characterize the popular view of "glassy" matter as irregular arrays of regular and irregular polyhedral (Miracle 2012), including metal glasses produced by supercooling (rapid quenching) or so-called "splat" cooling where liquid metals or other elements such as boron, or alloys such as 80Fe–20B (weight percent), are sprayed onto liquid nitrogen drum surfaces to produce amorphous ribbon. Since their discovery in 1960 (Klements et al. 1960), and especially magnetic Fe_xB (often called Metglass), more than a million transformer cores have been fabricated from splat ribbon laminates (Sheng et al. 2012; Staudhammer 2007).

Amorphous materials are metastable and will crystallize at what is called the glass transition temperature. For 80Fe–20B, the glass transition is around 470 °C. The recrystallization of amorphous boron creates crystalline α-B_{12}, which under conditions of elevated temperatures and pressure will transform to β-B_{106}. It is also observed in the amorphous boron structure shown in ▶ Fig. 22 of chapter "Electromagnetic Color and Color in Materials," and often amorphous silica glass structures, that individual atoms or smaller atom clusters can occupy the interstitial volumes and contribute to stabilizing (or "gluing") the structure. This is often difficult to establish because of the limitations of atomic level diagnostics.

Structures and Metallurgy of Plutonium

Having dealt with boron (Z = 5) and a few of its key compounds, we move to the end of the periodic table to plutonium (Z = 94) (shaded in ▶ Fig. 4 of chapter "Crystallography Principles"). Like boron, plutonium is polymorphic with six phases or crystal structures. It is far more dense than boron (~2.3 g/cm^3) ranging from 15.9 g/cm^3 for the fcc δ-phase to 19.9 g/cm^3 for the α-phase. Table 2 lists the plutonium phases (crystal structures and structure details) and stability ranges, while Fig. 9a illustrates the corresponding volume changes from the α-phase to the δ-phase, which in the aggregate is more than 25 %. Figure 9b also shows the extreme in tensile properties between α-Pu and δ-Pu in contrast to other common metals, bcc-Fe and fcc-Al, respectively. In some respects freshly cast or machined Pu looks somewhat like aluminum, but it dulls or tarnishes by oxidation in the air.

Table 2 Crystal structure data for plutonium[a]

Phase	Stability range °C	Space lattice and space group	Unit cell dimensions, A	Atoms per unit cell	X-ray density, gm/cm^3
α Alpha	Below ~115	Simple monoclinic $P2_1/m$	**@21 °C:** a = 6.183 ± 0.001 b = 4.822 ± 0.001 c = 10.963 ± 0.001 β = 101.79° ± 0.01°	16	19.86
β Beta	~115–~200	Body-centered monoclinic $I2/m$	**@190 °C:** a = 9.284 ± 0.003 b = 10.463 ± 0.004 c = 7.859 ± 0.003 β = 92.13° ± 0.03°	34	17.70
γ Gamma	~200–310	Face-centered orthorhombic $Fddd$	**@235 °C:** a = 3.159 ± 0.001 b = 5.768 ± 0.001 c = 10.162 ± 0.002	8	17.14
δ Delta	310–452	Face-centered cubic, $Fm\bar{3}m$	**@320 °C:** a = 4.6371 ± 0.004	5	15.92
δ′	452–480	Body-centered tetragonal $I4/mmm$	**@465 °C:** a = 3.34 ± 0.01 c = 4.44 ± 0.04	2	16.00
ε	480–640	Body-centered cubic, $Im\bar{3}m$	**@490 °C:** a = 3.6361 ± 0.004	2	16.51

[a]From Staudhammer (2007)

Table 3 shows the elastic (or Young's) modulus and shear modulus along with Poisson's ratio for the major Pu crystal structures.

Because of the ductility and stability of the fcc (δ)-phase, along with its lower density, δ-Pu is favored for weapons applications, which represents the only tractable application for Pu or its alloys. In general, Pu is extremely sensitive to even additions of alloying elements. Gallium is especially notable because it produces a broad region of stable δ-phase up to about 540 °C as illustrated in Fig. 10. Other elements such as Al, Zr, Ce, Th, Sc, and Zn have also been alloyed with Pu to stabilize the δ-phase. Figure 11 shows the typical polycrystalline grain structures for homogenized δ-Pu (Fig. 11b) and δ-Pu–Ga (~7 at.%) (Fig. 11c) in contrast to bcc tantalum (Z = 73) (Fig. 11a), having a nearly corresponding density of 16.6 g-cm^3.

On considering the huge range of polymorphic stability implicit in Table 2 and Figs. 9 and 11, it can be appreciated that early atomic weapons, because of their overall design and fissionable Pu–alloy compositions, could pose serious issues affecting reliability and associated stability. Even in the liquid–solid transition, Pu does not act like most other metals or alloys since it expands like the freezing of water. As noted above its density and volume change can exceed 25 %. In addition, Pu and its alloys react vigorously with water and its components, oxygen and hydrogen. Correspondingly it is pyrophoric at elevated temperatures in the

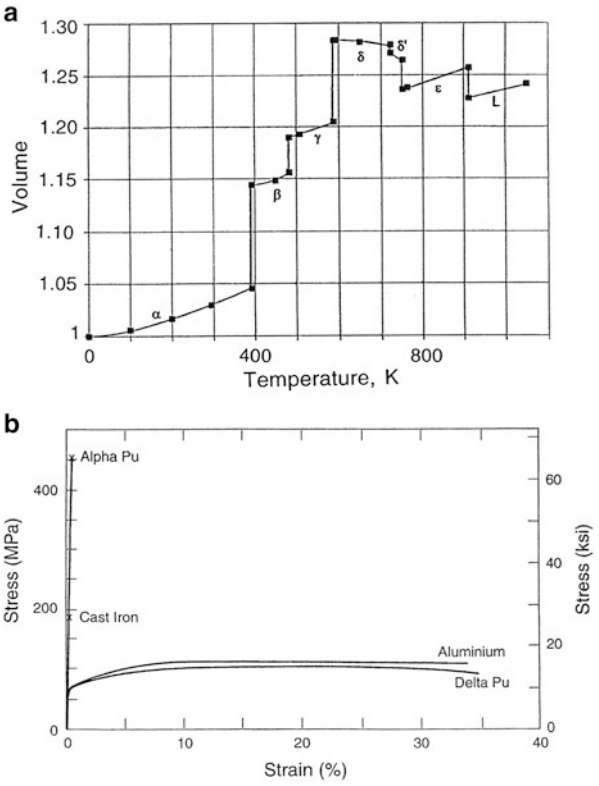


Fig. 9 Plutonium phase properties. (**a**) Composite approximation of Pu volume (thermal/expansion and phase transformation) versus temperature (From Schonfeld and Tate (1996)). (**b**) Stress–strain comparisons for α and δ plutonium and Fe and Al at room temperature (From Staudhammer 2007; Hecker 2000)

Table 3 Elastic constants for unalloyed Pu as a function of temperature

Temperature °C	Phase	Modulus of elasticity 10^6 psi	Shear modulus 10^6 psi	Poisson's ratio
−170	Alpha	17.35	7.26	0.20
−80	Alpha	16.00	6.69	0.20
0	Alpha	14.65	6.17	0.20
20	Alpha	14.00	6.08	0.16
60	Alpha	13.30	5.80	0.15
100	Alpha	12.40	5.45	0.14
140	Beta	5.90	2.29	0.29
170	Beta	5.58	2.20	0.27
200	Beta	5.37	2.12	0.27
220	Gamma	5.28	2.13	0.24
260	Gamma	4.94	1.97	0.25
300	Gamma	4.58	1.52	0.25
330	Delta	2.49	0.87	0.43
380	Delta	2.45	0.84	0.45
430	Delta	2.39	0.81	0.42

Note that 10^6 psi = 6.9GPa (Data from Staudhammer (2007))

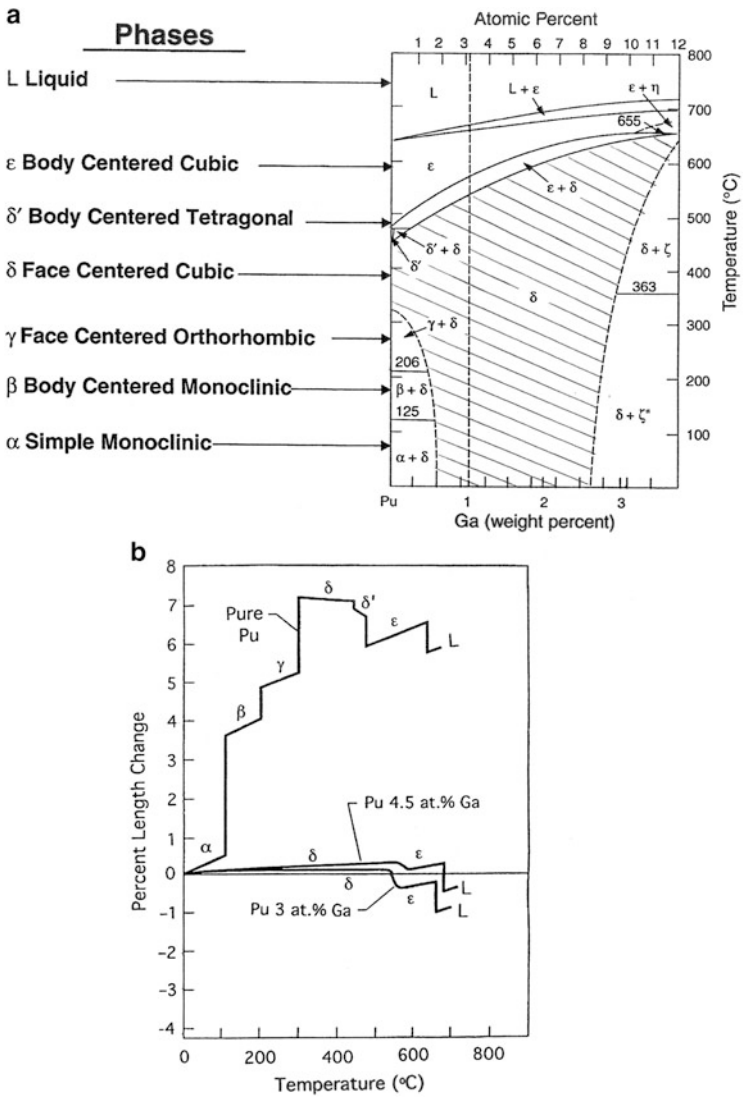


Fig. 10 Effect of Ga additions to Pu. (**a**) Pu–Ga phase behavior. (**b**) Thermal expansion for select Pu–Ga alloys (From Staudhammer (2007) and Hecker (2000))

atmosphere. As a liquid, Pu has the largest viscosity of any pure liquid metal and a surface tension of 550 mJ/m^2 in contrast to 720 and 2,150 mJ/m^2 for liquid Ga and Ta, respectively.

Thermally activated processes within Pu can change its internal structure with aging, and its continuous radioactive decay causes self-radiation damage such as the creation of vacancy–interstitial pairs (one radioactive decay event produces

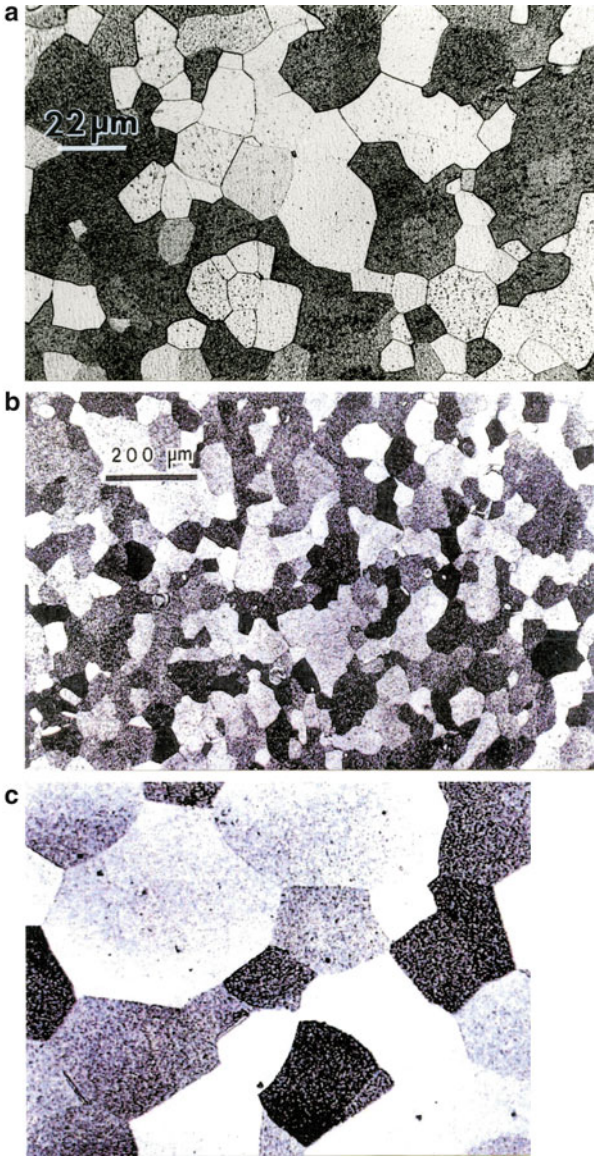


Fig. 11 Optical micrographs showing equiaxed grain structure for bcc tantalum (**a**), homogenized d-Pu (**b**), and homogenized (d) Pu–Ga alloy (c). The magnification in (c) is the same as (b) ((b) and (c) are from Staudhammer (2007))

around 2,500 such pairs) and the long-term formation of voids and void swelling. α-particle decay (emitting an α-particle or He^{2+}) created a uranium nucleus. The α-particles (He^{2+}) pick up electrons to become helium atoms which augment the void formation by aggregating to form bubbles. Each decay event releases ~5 million electron volts of energy which is converted to heat, making plutonium warm to the touch. There are 15 isotopes (Pu^{232} to Pu^{246}) with radioactive half lives ranging from 30 min for Pu^{233} to 8×10^7 years for Pu^{244}. After radiation release

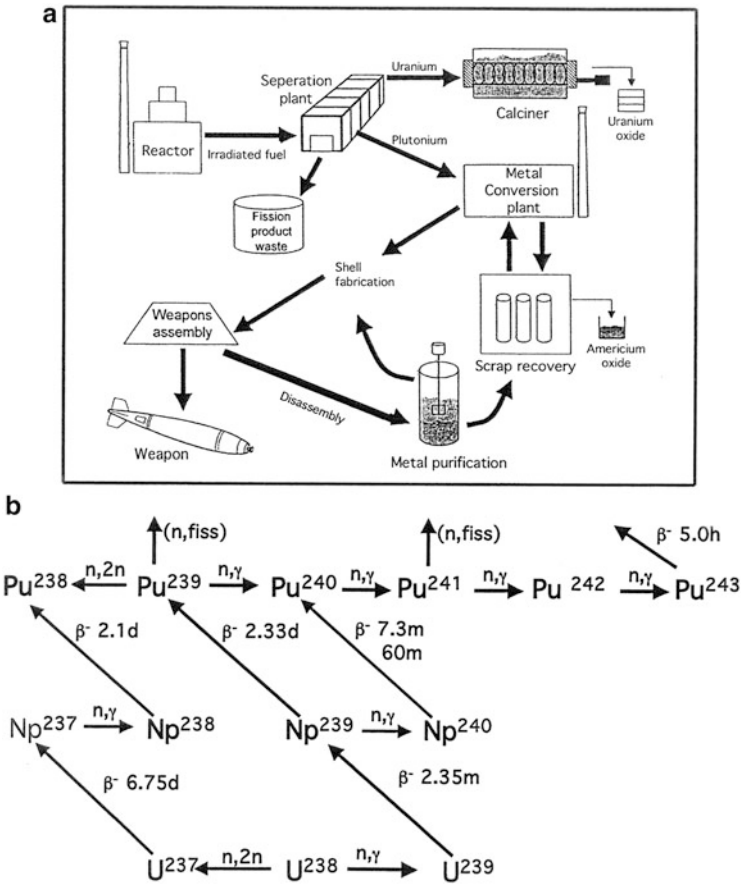


Fig. 12 Plutonium process cycle (**a**) and plutonium production from neutron irradiation of uranium in a nuclear reactor (irradiated fuel) (**b**). (**a**) is after Staudhammer (2007). (**b**) is adapted from Wick (1980)

(decay or release of a particle or energy) a new element is usually formed which is referred to as a daughter product in contrast to the decaying parent material. This decay process can also produce heat which for Pu isotopes ranges from about 560 W/h to 0.1 W/h for Pu^{242}. Pu^{238} oxide (PuO_2) glows red and is a liquid at ambient conditions; it can act as a long-term heat source (such as space power systems), although correspondingly radioactive. In addition to PuO_2, plutonium forms other compounds. Carbides and oxides find uses in cladding and pellet production.

Because of these features, nuclear weapons have been dismantled and reprocessed over the past decades. Reprocessed, Pu–alloy fissionable cores (or pits) have been recast, re-machined, and reassembled into new weapons. While roughly 100 metric tons of Pu have been processed in the USA over the

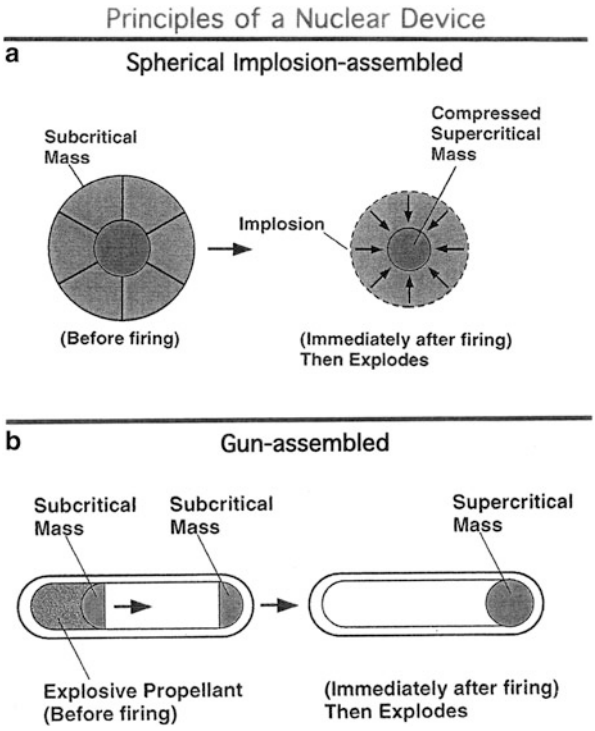


Fig. 13 Schematic views for triggering sustained fission in nuclear devices. (**a**) Spherical implosion of Pu (alloy) shell. Only eight detonation points are shown in the 2D projection view. In early nuclear devices, complex arrays of more than 60 detonation points were placed in the outer spherical explosive encapsulation. (**b**) Gun-driven hemi-shells create supercritical mass (From Staudhammer 2007)

years since 1943, the past decade or more (which has included weapon reductions by the USA and Russia in particular) has allowed new weapon production solely from processed or recycled Pu.

Figure 12a illustrates schematically the original production and processing of Pu, while Fig. 12b illustrates the production of plutonium from neutron irradiation of uranium in a nuclear reactor. The intermediate product is neptunium ($Z = 93$) between U ($Z = 92$) and Pu ($Z = 94$). Fissionable plutonium is observed to be Pu^{239} and Pu^{241}. Pu^{239} is alloyed (Fig. 10a) and configured to achieve a critical mass for a sustained nuclear reaction (fission) as illustrated schematically in Fig. 13. In Fig. 13a, a subcritical, spherical Pu–alloy shell is surrounded by high explosive and simultaneously detonated. The detonation sequence must be designed to implode the shell mass with high precision. Alternatively, a shell half is propelled against another shell half with similar explosive precision to create the compressed reaction which causes the nuclear explosion. Figure 14 illustrates the simultaneous detonation of a thin disk of C-4 ((RDX/91-$C_3H_6N_6O_6$) having a detonation velocity of $\sim$ 8 km/s) – a 2D implosion illustrating the precision of converging detonation waves recorded by a fast-framing camera over a period of roughly $2\mu s$/frame (or $\sim 30\mu s$). Figure 15 shows the actual test firing of a low-yield nuclear bomb in the USA prior to the cessation of atmospheric testing in 1964.

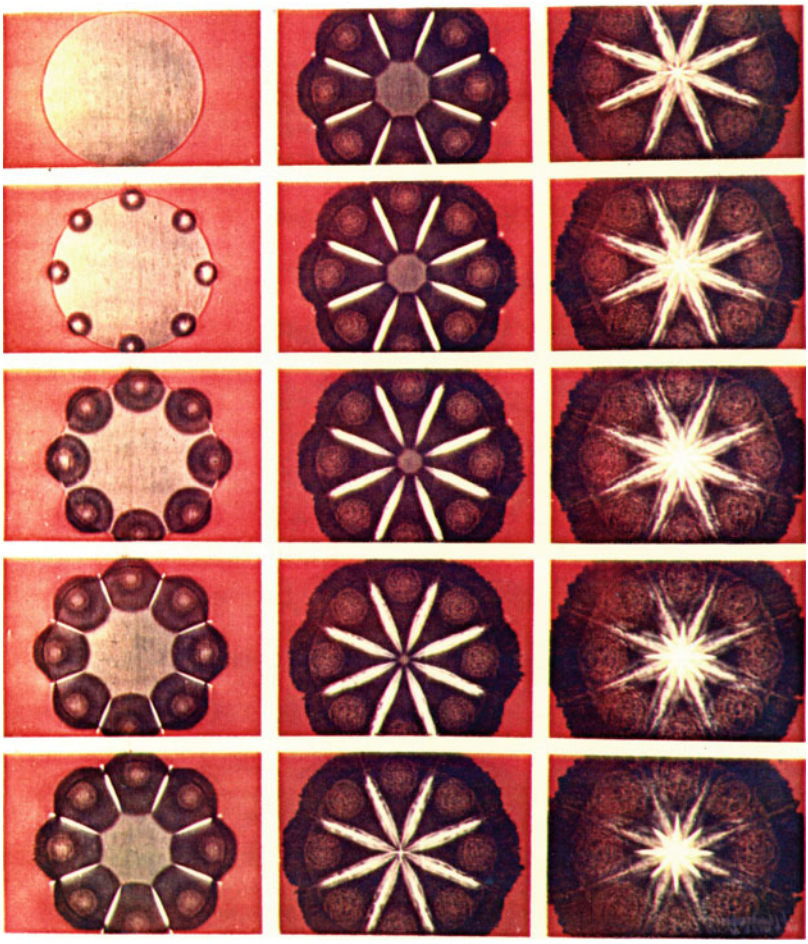


Fig. 14 Fast-framing camera sequence (~2μs intervals) for an 8-point detonation of a C-4 circular sheet 15 cm in diameter and 3 mm thick showing detonation wave convergence

The critical mass is the smallest amount of material required to sustain a fission chain reaction. Although the implications in Fig. 13 focus on the Pu core mass, other criticality parameters are involved, including geometry (Fig. 13b), volume, etc. Critical masses for δ-Pu (or alloys) generally range from ~6 to 16 kg.

Like polymorphic boron, polymorphic plutonium provides wildly contrasting materials properties which are narrowly applicable. For example, boron melts at 2,076 °C in contrast to plutonium which melts at 664 °C. Boron in its range of polymorphs (including amorphous B) has diamond pyramid microindentation hardnesses (HV) ranging from about 38 to 58 GPa in contrast to Pu with a microindentation hardness range of 0.5 (α)–5 GPa (δ). The elastic modulus for

Fig. 15 Detonation of nuclear device at a Nevada test site in the USA prior to nuclear test ban (From Staudhammer (2007))

boron ranges from 400 to 490 GPa (Table 1), while in contrast it ranges from 17 to 120 GPa in plutonium (Table 3). Boron crystal structures involve icosahedral, B_{12} building blocks forming complex bonding phenomena largely responsible for the high hardness and elastic constants. In contrast, plutonium forms classical metallic bonding in its polymorphs, and its properties emulate those of many common metals (Fig. 9b).

Plutonium, on the other hand, has had an enormous impact on global affairs, affecting most of the Earth's people. Its application in weapons of mass destruction imposes a certain level of stabilization as well as potential destabilization. It also poses some dangers in connection with its toxicity as a radiological hazard (Hecker 2000; Sutcliffe et al. 1995). Because it emits alpha particles, it delivers a negligible radiation dose to human skin. A few hundred grams in a glass jar can be held in the hand (covered with a rubber glove). However, when inhaled or ingested, the hazards are significant and can include cancer in various forms. It has been determined that 0.08 mg of inhaled Pu will have 100 % probability to cause cancer (Sutcliffe et al. 1995). However, in aerosol or nanoparticulate and respirable form, this would involve a relatively high concentration or nanoparticulate fraction. Because of the propensity for Pu to oxidize, the prospects for inhaling PuO_2 nanoparticulates are much more probabilistic, although the health risk is not abated in contrast to Pu (Wick 1980).

Although posing health risks, other radionuclides (or transuranic elements in the *actinide* series: elements from $Z = 89$ to $Z = 103$ in the periodic table of ▶ Fig. 4 of chapter "Crystallography Principles") have important applications in radiomedicine. These include neptunium (Np), americium (Am), and cerium (Cm), just before and after Pu ($Z = 94$). Neptunium production is illustrated in Fig. 12b and the Am–Pu, Np–Pu, and Cm–Pu phase diagrams are well established (Massalski 1990).

References

Barrett CS, Massalski TB (1966) Structure of metals, Third Edition, McGraw-Hill, New York.
Hecker SS (2000) Plutonium and its alloys. Los Alamos Sci 26:290–336
Hoard JL, Hughes RE (1967) In: Muetterties EL (ed) The chemistry of boron and its compounds. Elemental boron and compounds of high boron content: structure, properties and polymorphism. Wiley, New York, p 25
Hume-Rothery W (1936) The structure of metals and alloys. Institute of Metals, London
Klements W, Willen RH, Duwez P (1960) Non-crystalline structure in solidified gold silicon alloys. Nature 187(4740):869–890
Massalski TB (ed) (1990) Binary alloy phase diagrams, 2nd edn. ASM International, Metals Park
Miracle DB (2012) A physical model for metallic glass structures: an introduction and update. JOM 64(7):846–855
Schonfeld FW, Tate RE (1996) The thermal expansion behavior of unalloyed plutonium. Los Alamos National Laboratory, Los Alamos, LA-13034-MS
Sheng HW, Ma E, Kramer MJ (2012) Relating dynamic properties to atomic structure in metallic glasses. JOM 64(7):856–881
Staudhammer KP (2007) Introduction to plutonium metallurgy (ASM Los Alamos chapter short course). Los Alamos National Laboratory, Los Alamos, LA-UR-07-2485
Sutcliffe WG, Condit RH, Mansfield WG, Myers DS, Layton DW, Murphy PW (1995) A perspective on the dangers of plutonium. Lawrence Livermore National Laboratory, Livermore, UCRL-JC-118825
Wick OJ (ed) (1980) Plutonium handbook: a guide to the technology, vol I and II. The American Nuclear Society, La Grange Park
Widom M, Mihalkovic M (2008) Symmetry-broken bond crystal structure of elemental boron at low temperature. Phys Rev B77:064113-1-8

Aperiodic Crystal Structures: Quasicrystals

Contents

Abstract

It is remarkable that while the Platonic solids reflect 5-fold symmetry in both the icosahedra and the dodecahedron, crystal structures did not represent this symmetry until around 1983 when quasicrystals were discovered. This chapter presents a short overview of quasicrystals and 5-fold symmetric structures which have been described over the past three decades. Their relationship to the golden mean or the divine proportion is also briefly discussed. The concept of aperiodic or quasicrystals versus periodic or regular crystals is also described.

Introduction

As summarized in ▶ Figs. 1 and ▶ 2, and ▶ Table 1 of chapter "Crystallography Principles", periodic crystals are defined in three dimensions as periodic crystals of atoms with transitional periodicity along their three principal axes: x, y, and z (or a, b, c). They are described by 230 space groups designating rotational and translational symmetry elements corresponding to variances of the Bravais lattices (▶ Fig. 2 of chapter "Crystallography Principles") or crystal unit cells shown in ▶ Fig. 2 of chapter "Crystallography Principles". A quasicrystal in contrast to a regular (ordered and periodic) crystal characterized by a Bravais or modified Bravais lattice (▶ Fig. 2 of chapter "Crystallography Principles") is ordered but not periodic: aperiodic. While periodic crystals possess 2-, 3-, 4-, and 6-fold rotational symmetries (▶ Table 1 of chapter "Crystallography Principles"),

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_12

quasicrystals possess 5- or 10-fold rotational symmetries and no translational symmetry, often considered the key property characterizing the crystalline state. However a better characterization often considers long-range order (LRO) in contrast to short-range order (SRO) of the atomic structure. In LRO structures, single unit cells can be systematically stacked or translated as illustrated in ▶ Fig. 1 of chapter "Crystallography Principles".

Quasicrystals in the context of contemporary materials science and engineering were discovered by Dan Shechtman in the early 1980s and documented in a paper published with colleagues in 1984 (Shechtman et al. 1984), describing a rapidly solidified (splat cooled) Al_6 Mn alloy exhibiting icosahedral symmetry in electron diffraction patterns. The subsequent development of research on icosahedral materials led to a change in the definition of crystals as noted above to "material whose diffractogram has Bragg peaks" (International Union of Crystallography 1992). The award of the Nobel Prize in chemistry to Dan Shechtman in 2011 recognized the importance of this field, which by 2011 recorded nearly 10,000 literature publications.

The concept of quasicrystals was described in the decades prior to 1980 in relationship to incommensurate crystal structures, often referred to as "peculiar quasicrystalline materials." Incommensurate crystal structures lacked periodic translational symmetry not in an amorphous form, but because two or more elements of translational symmetry present in the structure were mutually incompatible. These structures were periodic structures modulated by a spatially periodic perturbation having a different length scale and often resembled "composites" of regular or SRO structures in an amorphous regime (or matrix) including 5- and 10-fold irrational symmetries (Riste 1980; Van Smaalen 2002). In this respect quasicrystals are ordered but not periodic, while amorphous solids are neither ordered nor periodic, in contrast to crystals which of course are both ordered and periodic.

This suggests that the stability of quasicrystals may lie in the realm between the amorphous state and the crystalline state. Indeed, quasicrystals were first prepared by splat cooling or the projection of a liquid stream onto a very cold surface to produce a rapidly solidified thin film (Riste 1980; Shechtman and Blech 1985), identical to the production of thin amorphous films (Klement et al. 1960; Hwang et al. 2012). In the context of growth phenomena, there are three types of quasicrystals: (1) stable quasicrystals (QCs) grown by slow cooling and anneal, (2) metastable QCs grown by melt spinning (splat cooling), and (3) metastable QCs grown in crystallization of an amorphous phase. As illustrated in ▶ Fig. 9 of chapter "Structure of Metals and Alloys", amorphous solids are often composed shells or clusters, and variations in the growth kinetics may favor their growth by shell additions as discussed in chapter "▶ Chemical Forces: Nanoparticles" or the aggregation of clusters into some space-filling regime to form irrational 3D structures. The difficulty in rationalizing these structures or quasicrystal analysis in determining these structures comprises the SRO of the atomic arrangement inside the clusters in contrast to the LRO of the clusters themselves.

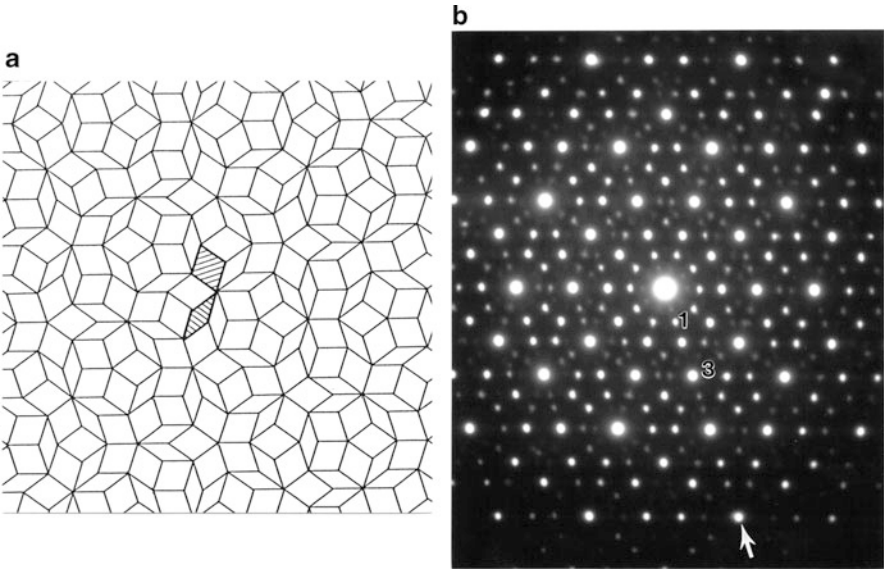


Fig. 1 Aperiodic 2D Penrose tiling (**a**) and 5-fold symmetric electron diffraction pattern for Al_{65} Cu_{22} Fe_{13} quasicrystal thin film (**b**)

Quasicrystal Structures

Since quasicrystals have lost periodicity in at least one dimension, they cannot be described in 3D space like regular crystals. Diffraction from crystals as shown, for example, in ▶ Fig. 8 of chapter "Lattice Directions and Planes, and Diffraction by Crystals" is characterized by the 3 miller indices which describe the 3D translational periodicity of the structure (▶ Fig. 6 of chapter "Lattice Directions and Planes, and Diffraction by Crystals"). Quasicrystals, on the other hand, have no 3D translational symmetry and their diffraction patterns would require 5 rather than 3 indices with a reciprocal space characterized by icosahedral symmetry or some higher-dimensional space.

Penrose (1974) among others (Steurer and Deloudi 2009) has demonstrated that it is possible to construct a 2D crystal with 5-fold rotational symmetry from a set of two rhombic tiles. This 2D tiling can also be generalized to 3D using appropriate rhombohedral unit cells based on these two rhombuses. Unfortunately this is not applicable in a general sense of periodic structures since the concept is in principle aperiodic.

Figure 1 provides an overview of practical aspects of quasicrystals we have described above. Figure 1a shows the two-dimensional Penrose tiling of the plane generated by appropriate stacking of the two shaded rhombuses (or rhombs). Corresponding to the 5-fold symmetry represented in this tiling, Fig. 1b reproduces an electron diffraction pattern for icosahedral Al_{65} Co_{20} Cu_{15} which is typical for

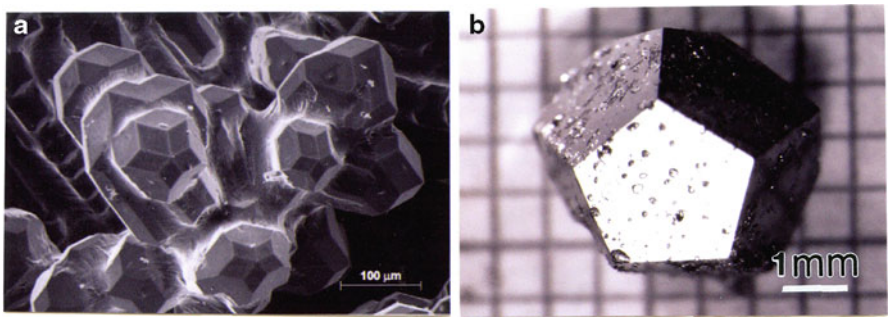


Fig. 2 Examples of quasicrystals. (**a**) Al-Pd-Re ternary quasicrystal (Courtesy of I. Fisher, Stanford University). (**b**) Al_{65} Cu_{22} Fe_{13} dodecahedral quasicrystal

essentially all 5- and 10-fold symmetric quasicrystal diffraction patterns. In addition, Fig. 2 shows examples of quasicrystals which illustrate these symmetry features. Figure 2a in fact shows a 3D tiling structure which replicated the Penrose tiling in Fig. 1a, while Fig. 2b shows a single-crystal dodecahedron measuring several millimeters, which is the dual of the icosahedron (▶ Fig. 1 of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements").

It can be noted that the two rhombs shown shaded in ▶ Fig. 30a of chapter "Electromagnetic Color and Color in Materials" forming the 2D Penrose tiling are based on the so-called golden mean: $\phi = 1.618$. Sometimes designated τ (and referred to as the divine proportion by Leonardo da Vinci):

$$\phi \equiv \tau = \frac{1+\sqrt{5}}{2} = 1.618 \tag{1}$$

which is also the root of the quadratic equation

$$\varphi^2 - \varphi - 1 = 0 \tag{2}$$

In this respect, an icosahedron can be inscribed in an octahedron so that each vertex of the icosahedron divides an edge of the octahedron in the ratio of ϕ:1. Correspondingly, the icosahedron is a derivative of the cuboctahedron as illustrated in ▶ Fig. 2 of chapter "Chemical Forces: Nanoparticles". It might also be noted that

$$\phi = \lim_{n\to\infty}\left[\frac{F(n+1)}{F(n)}\right] \tag{3}$$

where

$$F(n) = \frac{\phi^n - (1-\phi)^2}{\sqrt{5}} = \frac{\phi^n - (-\phi)^{-n}}{\sqrt{5}} \tag{4}$$

describes a one-dimensional sequence called a 1D Fibonacci lattice having the form
0, 1, 1, 2, 3, **5**, **8**, 13, 21, 34, **55**, **89**, . . .

Where the ratios F(n + 1)/F(n) converge to the golden mean; $\phi = 1.618$: 8/5 (above) = 1.6, while 89/55 = 1.6181818. It can be observed in the diffraction pattern of Fig. 1b that considering any of the 10 symmetry axes to be represented by a 1D Fibonacci lattice (arrow in Fig. 1b), the distances of successive diffraction spots divided by previous distances also equal the golden mean. Alternatively each bright spot from the origin (or center spot) can be considered a sequence beginning with the first spot labeled 1, the next (2) ϕ, the third spot, ϕ^2, the fourth, ϕ^3, and the fifth, ϕ^4 (located at the arrow in Fig. 1b), etc. Correspondingly, the side of a regular decagon inscribed in a unit circle is equal to $1/\phi$. This is also the base of the golden triangle and two such triangles arranged base-to-base form over of the golden rhombs of the Penrose tiling (Fig. 1a). The other rhomb has a long diagonal-to-short diagonal equal to ϕ.

The problem with quasicrystals is that clusters having many polyhedral or multi-polyhedral forms (multi-shell forms) dominate, as illustrated in Fig. 2. The occupying of periodic lattices or lattice points in 3D, such as illustrated for fullerite in ▶ Fig. 10 of chapter "Chemical Forces: Nanoparticles" or boron polymorphs shown in ▶ Fig. 5 of chapter "Structure of Metals and Alloys", are not quasicrystals since the clusters act like soft superatoms according to the *Jellium model* illustrated in ▶ Fig. 3 of chapter "Chemical Forces: Nanoparticles" and create periodic crystal structures having large unit cells (▶ Fig. 10 of chapter "Chemical Forces: Nanoparticles"). While a golden rhombohedron composed of six rhombs whose diagonals (long to short) are in the golden mean ratio (ϕ) can fill 3D space (illustrated ideally in ▶ Fig. 1b of chapter "Electromagnetic Color and Color in Materials"), quasicrystals do not assume this single structure. Furthermore, golden rhombs form faces of the rhombic triacontahedron illustrated in ▶ Fig. 1b of chapter "Chemical Forces: Nanoparticles", and while these polyhedra can fill space, they are also not completely representative of quasicrystal structure dominated by clusters and complex cluster structure.

As discussed in a comprehensive, 20-year review of quasicrystal structure research by Steurer (2004), quasicrystal phases with diffraction patterns resembling quasiperiodic structures (Fig. 1b) were found in roughly one hundred metallic binary and ternary systems. According to the observed diffraction symmetries, quasicrystals are classified as icosahedral (i), pentagonal (p), octagonal (o), decagonal (d), and dodecagonal (dd) QC phases. To date all quasicrystal structures can be described by polyhedral clusters and multi-shell clusters characterized in the Platonic solids shown in ▶ Fig. 1 of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements", as well as 13 Archimedean solids and 2 Keplerian solids shown in Fig. 3. Only a few of these can be packed in space-filling regimes as shown in Fig. 4. The packing of truncated octahedra shown in Fig. 4a is often referred to as ideal grain packing in polycrystalline cubic (fcc, bcc) metals and alloys where individual (periodic) crystal grains are characterized by the truncated octahedron (Fig. 3d).

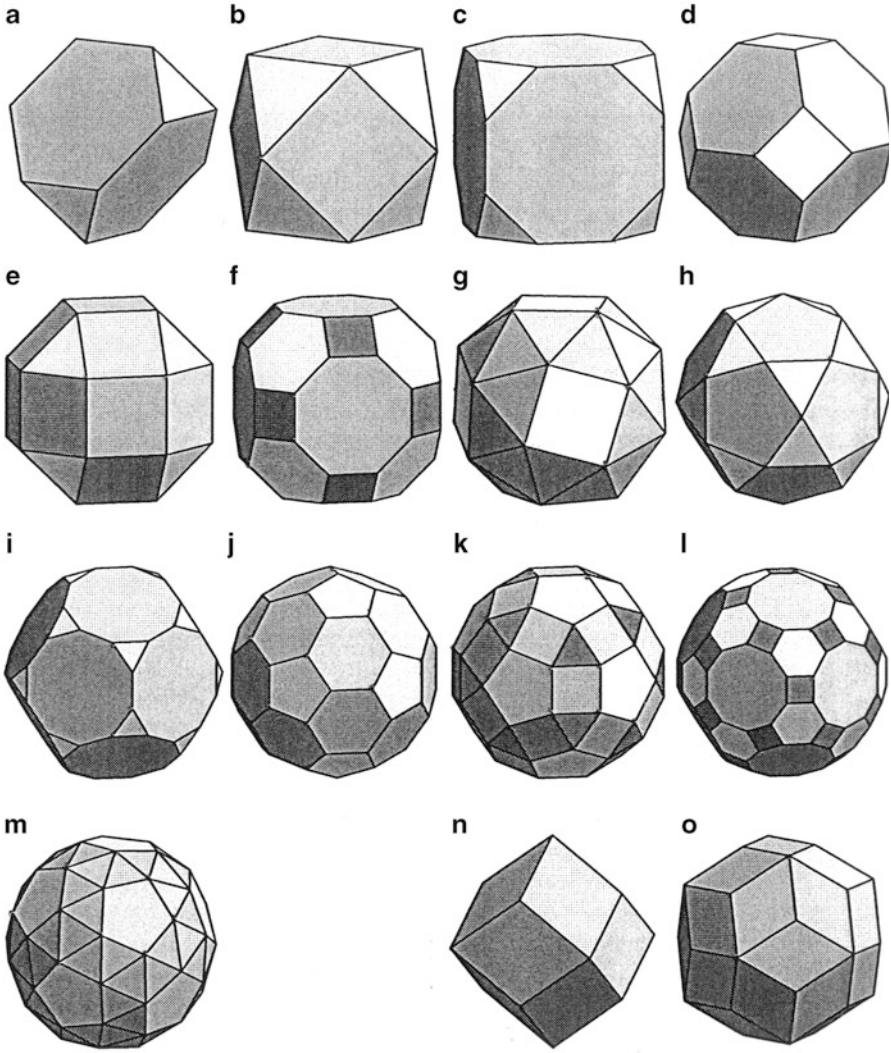


Fig. 3 The 13 Archimedean solids (**a**–**m**) and the Keplerian solids (**n**, **o**): (**a**) truncated tetrahedron, (**b**) cuboctahedron, (**c**) truncated cube, (**e**) (small) rhombicuboctahedron, (**f**) truncated cuboctahedron (great rhombicuboctahedron), (**d**) truncated octahedron, (**g**) snub cube, (**h**) icosidodecahedron, (**i**) truncated dodecahedron, (**j**) truncated icosahedron, (**k**) (small) rhombicosidodecahedron, (**m**) snub dodecahedron. The rhombic dodecahedron, (**n**), and the rhombic triacontahedron, (**o**), are duals of the cuboctahedron (**b**) and the icosidodecahedron (**h**) (Adapted from Steurer and Deloudi (2009))

The rhombic triacontahedron in Fig. 4c is often called an edge and face-transitive dual of the icosidodecahedron (Fig. 3h), composed of 30 golden rhombs joined at 60 edges and 32 vertices (twelve 5-fold and twenty 3-fold types). They can form 3D covering or cluster outer shells which can form overlapping rhombic

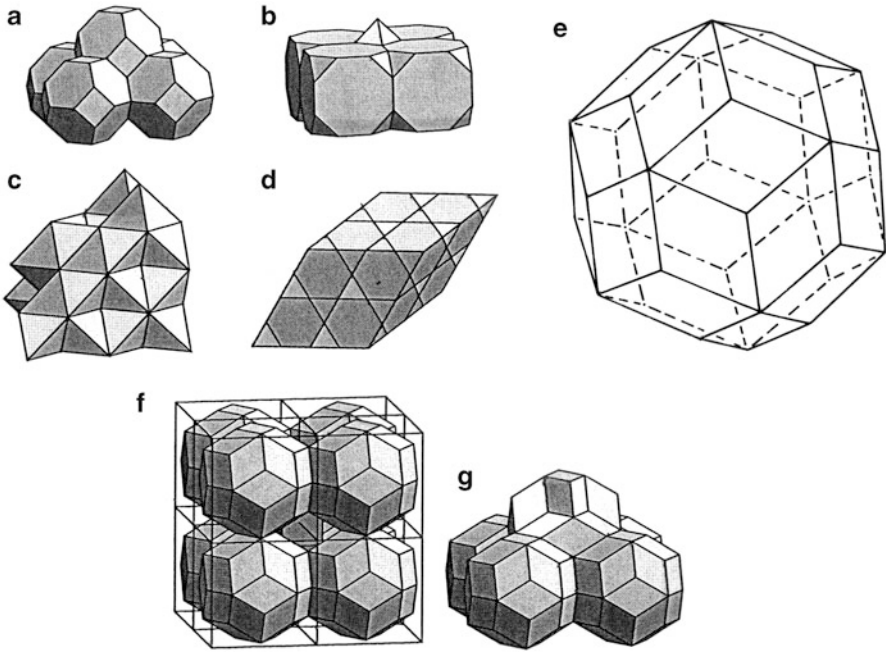


Fig. 4 Examples of packing and space filling of polyhedral. (**a**) Truncated octahedral, (**b**) truncated cubes + octahedra, (**c**) octahedra + tetrahedra, (**d**) truncated tetrahedra + tetrahedra, (**e**) rhombic triacontahedron, (**f**) packing of rhombic triacontahedra by sharing a face along each of the eight 2-fold directions, (**g**) the remaining empty space is a dimpled triacontahedron (with eight oblate rhombohedra removed) (Adapted from Steurer and Deloudi (2009))

triacontahedral (RTH) units as illustrated in Fig. 5. These can be built up of successive polyhedral shells shown previously in ▶ Figs. 1 of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements" and ▶ 3 of chapter "Aperiodic Crystal Structures: Quasicrystals". Only i-QCs have been shown to exhibit LRO and very few have been identified with reasonable accuracy. i-$YbCd_{5.7}$ (Takakura et al. 2007) is a recent example as illustrated in Fig. 6 which shows bcc packing of RTH units (92 atoms) filled with a Cd icosidodecahedron (30 atoms) (Fig. 3h), a Yb icosahedron (12 atoms), a dodecahedron (20 atoms), and an inner Cd tetrahedron (▶ Fig. 1 of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements") (4 atoms). The RTH unit can be considered to be composed of acute and obtuse rhombohedral (AR and OR) building units proposed by Henley (1991).

The i-$YbCd_{5.7}$ phase structure in Fig. 6 was analyzed using synchrotron XRD (Takakura et al. 2007). While high-resolution XRD may be used as a QC "fingerprint," it cannot be used as the sole proof for aperiodicity. More convincing representations of QC aperiodicity are 5-fold symmetric (or 10-fold symmetric)

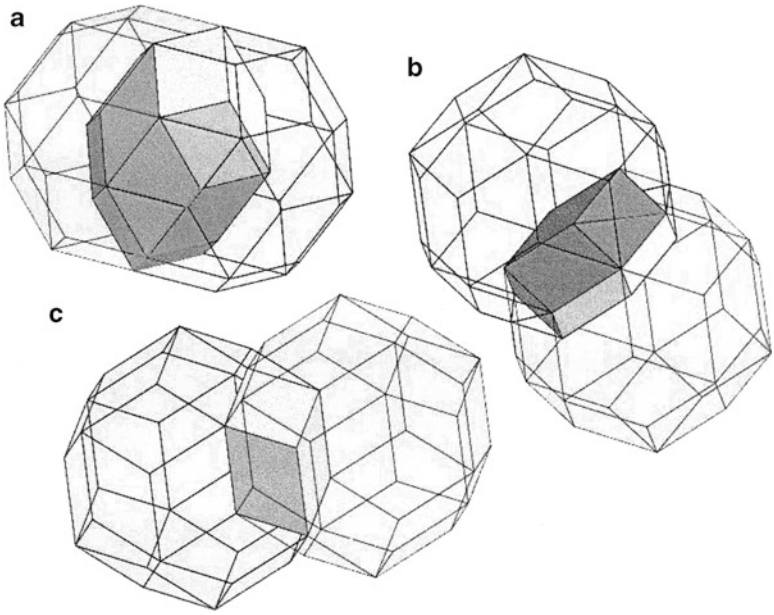


Fig. 5 Overlapping rhombic triacontahedra along the 5-, 3-, and 2-fold directions, (**a**–**c**) respectively. The shared volumes are characterized by a rhombic icosahedron (**a**), an oblate golden rhombohedron (**b**), and a rhombic dodecahedron (**c**)

electron diffraction patterns as shown typically in Fig. 1b, representing the original discovery by Shechtman et al. (1984) and Shechtman and Blech (1985).

Despite the validation of the fundamental importance of QCs implicit in their recognition by the award of a Nobel prize (in 2011 as noted above), their difficulty in producing useful forms of these materials limits their applications. Like boron polymorphs represented in ▶ Table 1 of chapter "Structure of Metals and Alloys", QCs have exhibited micro indentation hardnesses (HV) ranging from roughly 5 GPa (Al-Li-Cu; i-Ho-Mg-Zn) (Takeuchi et al. 1991) to as high as 22 GPa (i-Al-Pd-Mn) (Trebin 2006), 9 GPa (Al_3Co), 11 GPa (Al_5Co_2), and 10 GPa ($Al_{65}Cu_{20}Co_{15}$) (Murthy et al. 1999), among other measurements. These high hardnesses are related to the difficulty in moving dislocations in these aperiodic structures and suggest applications involving high-strength (bond) surface coatings formed by rapid solidification. Efforts to create useful metal-matrix composites using i-QC particles have not proven to be useful (Trebin 2006).

To date, there has been only one observation of a QC phase in a terrestrial mineral found in the Khatyrka river region of eastern Russia and determined to be $Al_{63}Cu_{24}Fe_{13}$ (Bindi et al. 2011). The origin of this geological sample may, however, be extraterrestrial in origin. Nonetheless, this observation may be an indication of QC stability.

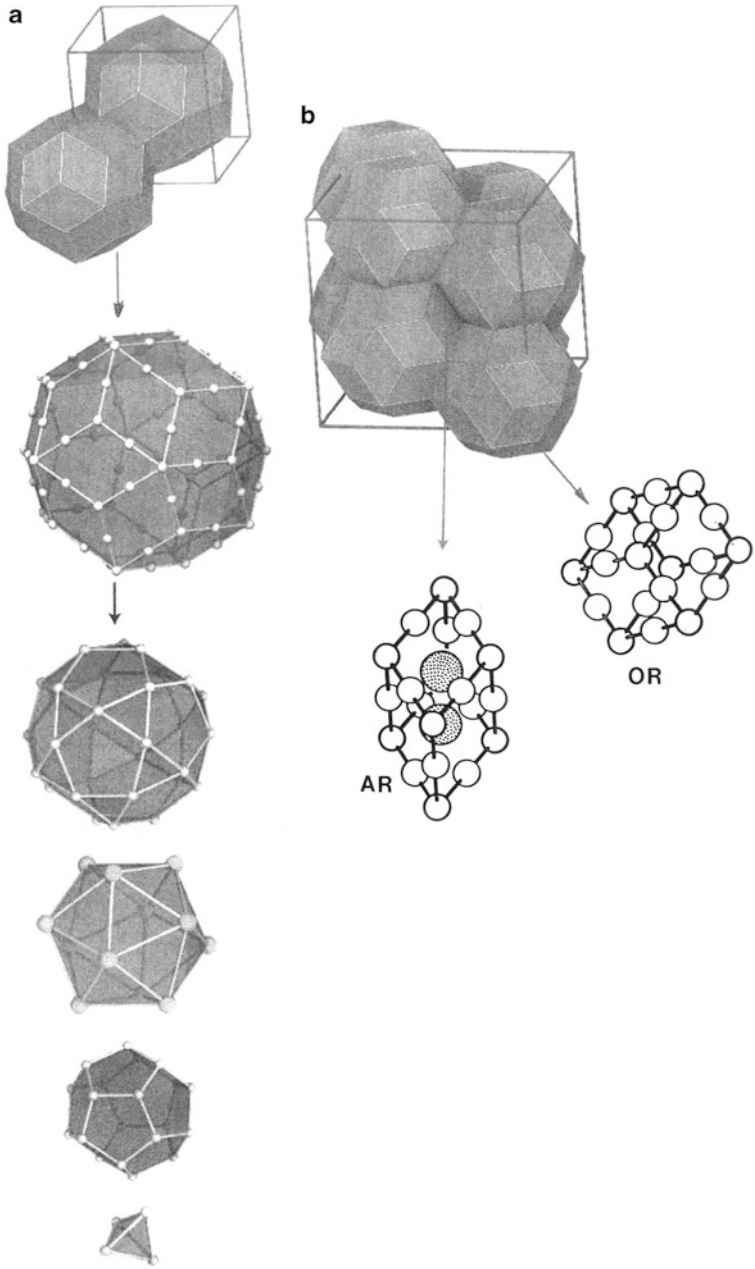


Fig. 6 $Y_{13}Cd_{76}(Yb\ Cd_{5.7})$, i-QC structure components. (**a**) bcc packing of RTH units in 1/1 approximant with fundamental polyhedral shell components. (**b**) shows a 2/1 approximant with acute rhombohedron (*AR*) and obtuse rhombohedron (*OR*) building units (Adapted from Takakura et al. (2007))

References

Bindi L, Steinhardt PJ, Yao N, Lu PJ (2011) Icosahedrite, $Al_{63}Cu_{24}Fe_{13}$, the first natural quasicrystal. Am Mineral 96:928–931

Henley CL (1991) Cell geometry for cluster-based quasicrystal models. Phys Rev B 43:993–1020

Hwang J, Melgarejo ZH, Kalay YE, Kalay I, Kramer MJ, Stone DS, Voyles PM (2012) Nanoscale structure and structural relaxation of $Zr_{50}Cu_{45}Al_5$ bulk metallic glass. Phys Rev Lett 108:195505

Klement W, Willens RH, Duwez P (1960) Non-crystalline structure in solidified gold silicon alloys. Nature 187(4740):869–870

Murthy GVS, Roy AK, Minz RK, Mukhopadhyay NK (1999) Microhardness and fracture toughness studies of decagonal quasicrystals in Al-Cu-Co system. J Mater Sci Lett 18(3):255–258

Penrose R (1974) The role of aesthetics in pure and applied mathematical research. Bull Inst Math Appl 10:266–271

Riste T (ed) (1980) Ordering in strongly fluctuating condensed matter systems. Plenum Press, New York

Shechtman D, Blech IA (1985) The microstructure of rapidly solidified Al_6Mn. Metall Trans 16A:1005–1012

Shechtman D, Blech I, Gratian D, Cahn J (1984) Metallic phase with long-range orientational order and no translational symmetry. Phys Rev Lett 53:1951–1953

Steurer W (2004) Twenty years of structure research on quasicrystals. Part I. Pentagonal, octagonal, decagonal and dodecagonal quasicrystals. Z Kristallogr 219:391–446

Steurer W, Deloudi S (2009) Crystallography of quasicrystals: concepts, methods and structures. Springer, New York

Takakura H, Gomez CP, Yamamoto A, deBoissieu M, Tsai AP (2007) Atomic structure of the binary icosahedral Yb-Cd quasicrystal. Nat Mater 6:58–63

Takeuchi S, Iwanaga H, Shibuya T (1991) Hardness of quasicrystals. Jpn J Appl Phys 30:561–562

Trebin H-R (ed) (2006) Quasicrystals: structure and physical properties. Wiley, New York

Van Smaalen S (2002) Incommensurate crystallography. Oxford University Press, Oxford

Electrovalent Crystal Structures and Chemistry

Contents

Abstract

Electrovalent or ionic crystals, especially oxide crystals, exhibit interesting properties as a consequence of vacancy-based stoichiometry variations and related phenomena such as color center production described in chapter "► Electromagnetic Color and Color in Materials". In this chapter, metal oxides and sulfides are treated as AX and AX_2 compounds where X represents oxygen or sulfur, as well as perovskite and spinel structure oxides iABO_3 and AB_2O_4 structures, respectively. Oxide structures and crystal chemistry are particularly interesting because they are precursors to oxide superconductors which are treated in chapter "► Structures and Properties of Oxide Superconductors" which follows.

Introduction

While compositions of metals, even simple binary mixtures of metals forming binary alloys allows their structures, properties, and performance to be selectively and often systematically manipulated, ionic crystals formed from combinations of metal and oxygen ions can be manipulated in similar ways. However, in ionic solids, change balance dictates the chemistry or stoichiometry and is inviolate. In the case of metal oxides, the oxygen valence is always −2. Consequently, mixed metal valence states or variations in cation (metal ion) composition may require

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_13

vacant cation sites. In transparent crystals, especially oxide crystals, vacant sites will alter the optical absorption spectrum. Similarly, vacant ionic or anion (oxygen) sites will similarly affect optical properties while as noted in chapter. "► Electromagnetic Color and Color in Materials," when anion sites are occupied by electrons, so-called color centers are created which also alter optical absorption, such as F-centers in NaCl or CaF_2, where electrons occupy either the Cl^- or F^- anion sites in the crystal structure (chapter "► Electromagnetic Color and Color in Materials"). In ionic crystals, both the charge and lattice site must be considered. For example, a vacant site carries no charge. Consequently in an AX compound where A is (+) and X is (−), a vacancy in the cation site (A^+) either requires a corresponding vacancy in an anion site (X^-) or a substitution of +2 in an adjoining cation site: $A^+\square B^{2+}(3X^-)$ 3 (cation sites): 3 (anion sites). This changes the chemistry of the crystal. However, the chemistry would be the same for A^+ and A^{2+}, but the composition or stoichiometry would change accordingly. Such valence-state variations can occur for A^{2+} and A^{3+} when X^{-2} characterizes the anion sites. For example, FeO ⇒ $Fe^{2+}O^{-2}$ while a mixed-valence cation state can occur for Fe_x^{2+} $Fe_y^{3}\square_z O^{-2}$ where 2x + 3y must equal the anion charge (−2): 2x + 3y = 2 and x + y + z = 1 cation site. This is explored in more detail in the following section.

AX Structures: Metal Oxides (*MO*) and Sulfides (*MS*)

As noted previously, there are thousands of simple binary alloys and compounds when considering the more useful and available elements of the periodic table (AB). There are specific categories of metal compounds having considerable structural, particularly crystal structure, flexibility on considering valence substitution ions in cation sites and charge balancing electrovalent (or ionic) crystal structures. Metal oxides are of particular importance because of their wide uses in many material systems and their ability to alter their properties through cation valence substitutions and associated crystal structure modifications, including color center phenomena discussed in connection with ► Figs. 10 and ► 11 in chapter "Electromagnetic Color and Color in Materials".

The most stable oxides involve stable divalent (2+) cation elements since oxygen is almost invariably a divalent (−2) anion element. Stable divalent metals are the second column (or group) of the periodic table shown in ► Fig. 6 in chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements": Be, Mg, Ca, Sr, Ba, and Ra (which is relatively rare). Transition metal oxides, excluding CaO, ScO, and ZnO, can form multivalent oxides: TiO → TiO_2 (Ti_nO_{2n-1}; n = 1, 2, …) and FeO → Fe_x^{2+} $Fe_y^{3+}O$; NiO → $Ni_x^{2+}Ni_y^{3+}O$. Many of these MO oxides and MS sulfides are NaCl-type structures while others are zincblende (ZnS) or wurtzite, or ZnS-type structures. CuO, AgO, PdO, and PtO follow the special PdO-type structure. These structure types are shown for comparison in Fig. 1, while Table 1 lists the crystal structure types for these metal oxides along with several corresponding sulfides.

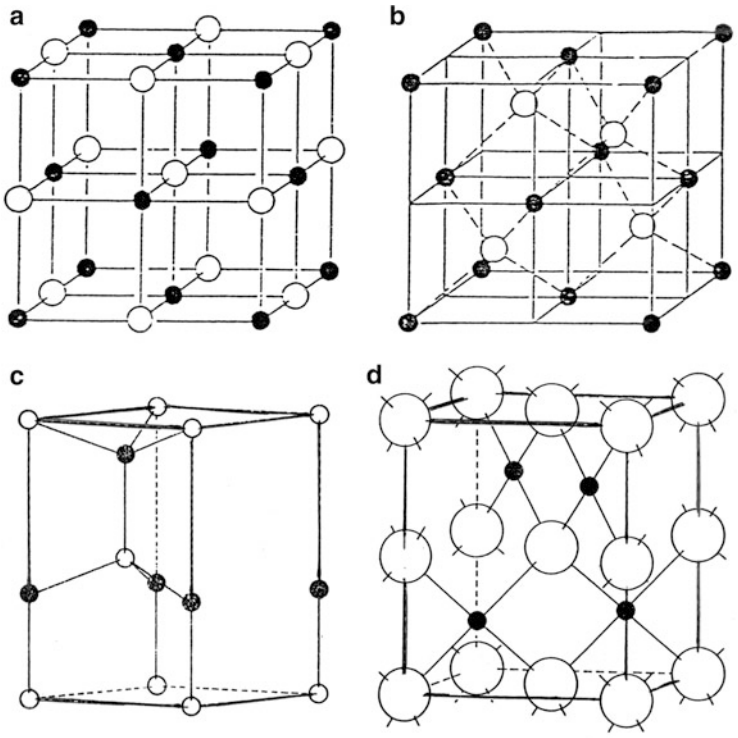


Fig. 1 MO and MS crystal structure types. (**a**) NaCl-type, (**b**) zincblende (ZnS) or diamond-cubic type, (**c**) Wurtzite (ZnS) hexagonal structure type, and (**d**) PdO tetragonal structure type

Table 1 Ax oxide and sulfide structure types

NaCl type		Zincblende (ZnS) type		Wurtzite (ZnS) type		PdO type	
Oxide	Sulfide	Oxide	Sulfide	Oxide	Sulfide	Oxide	Sulfide
MgO	MgS	BeO	BeS	ZnO	MnS	CuO	PdS
CaO	CaS		ZnS		ZnS	PdO	PtS
SrO	SrS		CdS		CdS	PtO	
BaO	BaS		HgS			AgO	
ScO							
TiO							
VO							
MnO							
FeO							
CoO							
NiO							

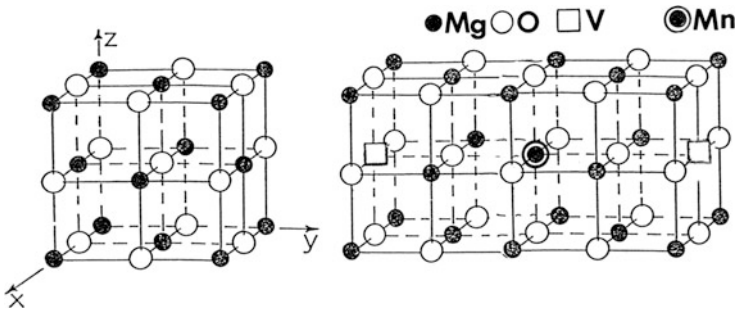


Fig. 2 $Mg_6\,Mn\square O_8$ ordered solid solution. Cation site vacancy is designated □(V). Based on the cubic NaCl structure shown at the *left*. The structure at the *right* represents 2 NaCl structure types: 2 × (4MgO)

It might be useful to point out that like most circumstances regarding crystal structures, in particular, the assumption of specific structures by AX compounds is driven by issues contributing to minimum energy configurations. The most obvious in examining Fig. 1 is related to coordination and the distances between cations and anions which are related to their respective sizes or the cation/anion radius ratio: r^+/r^-. The potential energy decreases with a decrease in the cation-anion distance, and the crystal structure becomes more stable. In the case of the NaCl-type structure as it is rendered in Fig. 2, if the smaller-size atoms are cations, as these decrease in size, the anion-anion distance decreases until they touch. This will occur at a radius ratio: $r^+/r^- = 0.414$ (or $\sqrt{2}-1$). For the zincblende structure shown in Fig. 1b, when considered on the same geometrical basis, the critical radius ratio corresponding to anion-anion contact will be given by $r^+/r^- = 0.414$ (or $\sqrt{6}/2-1) = 0.225$, and the corresponding energy of this structure is much greater than the NaCl-type structure in Fig. 1a.

It can be observed that the NaCl-type AX compounds in Table 1 are primarily ionic while for ZnS types (which are polymorphs for several sulfides), the bonding is primarily covalent, although BeO maintains a significant ionicity. However, the tetragonal PdO-type structure is a predominantly covalent bonding structure type. Generally in structures other than the NaCl structure, sulfides tend to be more covalent bonding than oxides.

There are many examples of solid solution ordering and cation vacancy formation to compensate for cation valence-elements substitutions. MgO provides an example:

$$\left(Mg_6^{2+}Mn^{4+}\square^0\right)O_8^{-2}\ ;$$

which can be illustrated in two adjoining NaCl-type crystal structures shown in Fig. 2 since 4MgO =1NaCl structure unit cell: 4 Mg fcc and 4 O fcc. This structure (Fig. 2) is charge balanced ($Mg_6^{2+} + Mn^{4+} = +12 + (+4) = +16$; $\square^0$; vacancy having no mass and no charge occupies one Mg (cation) site: +16 charge and 8 oxygen (anion) sites).

A similar ordered structure would be

$$\left(Cu_6^{2+}Pd^{4+}\square^0\right)O_8^{-2}$$

which would appear like Fig. 2. Consider the case of NiO → $(Ni)_{0.9}O : (Ni_x^{2+}Ni_y^{3+})_{0.9}\square^0_{0.1}O^{-2}$.

To balance the structure:

Charge balance: 2x + 3y = +2
Cation site balance: x + y = 0.9
Solving: y = 0.2; x = 0.7

NiO is a green (colored) oxide which is nonconducting (insulating) while $(Ni)_{0.9}O \rightarrow (Ni^{2+}_{0.7}\ Ni^{3+}_{0.2}\ \square^0_{0.1})\ O^{-2}$ is a black oxide conductor (or semiconductor). This stoichiometry can be rationalized in 10 (NiO) molecules: $(7\ Ni^{2+}2Ni^{3+}\ 1^0)\ 10\ O^{-2}$, representing 10 cation sites and 10 anion sites, which can be shown in two and one-half NaCl-type crystal structure units in contrast to the two-NaCl-type crystal structure units illustrated in Fig. 2 for 8 MgO.

AX_2 Structures: MO_2 and Related Oxides

There are a number of important metal oxides, MO_2, where either the fluorite (CaF_2)-type (cubic) structure or the rutile-type (tetragonal) structure dominates as illustrated in Fig. 3. There are no prominent sulfides (MS_2) assuming these structure types. TiO_2 is the prominent rutile structured dioxide and represents the end point of the oxidation states for titanium which ideally begins with TiO in the NaCl-type (cubic) structure (a = 4.18 Å), and as TiO_2, the crystal structure is tetragonal rutile (a = 4.59 Å, c = 2.96 Å; S. G. 4 mmm). This progression of oxidation states can be represented by Ti_nO_{2n-1}: TiO, Ti_2O_3, Ti_3O_5, Ti_4O_7. . .TiO_2 where beyond TiO, the valence is mixed, Ti^{2+}, Ti^{4+}, until TiO_2 where it is Ti^{4+}: $Ti_2O_3 \equiv Ti^{2+}Ti^{4+}O_3^{-2}$; $Ti_3O_5 \equiv Ti^{2+}Ti_2^{4+}O_5^{-2}$; etc. In addition, like NiO illustrated previously, TiO (NaCl-type structure) is usually non-stoichiometric and can vary from $Ti_{0.8}O$ to ideally TiO. For $TiO_{0.8}O$: $Ti^{2+}_{0.6}Ti^{4+}_{0.2}\square^0_{0.2}O^{-2}$, which can be ordered, while other stoichiometries may not be ordered. These oxidation states for titanium oxide structures, together with vacancy concentrations, which may be ordered or disordered, can alter the electronic structures just as in Ni_xO structures discussed previously.

It might also be noted that TiO_2 has 6 polymorphs including tetragonal rutile. Anatase (tetragonal) and brookite (orthorhombic) are prominent. TiO_2 as rutile is very white in powder or nanopowder form and is the base for white paint.

There are similarities in the M_OO_2-to-MoO3 oxidation states where we can have the following:

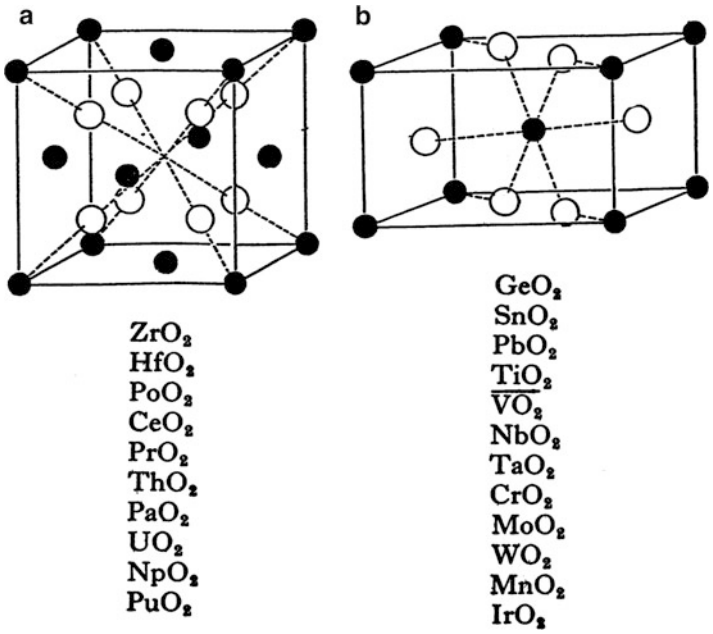


Fig. 3 (**a**) Fluorite (CaF_2)-type structure and (**b**) rutile-type (tetragonal) structure and their representative MO_2 oxides. Metal atoms (or ions) are solid circles while oxygen is open circles

Mo_nO_{3n-1}; MoO_2 (red)
Mo_4O_{11} (wine red)
Mo_5O_{14} (violet)
Mo_8O_{23} (violet)
MoO_3 (green yellow)

which illustrates a few examples of molybdenum oxides and their color variations which are also somewhat characteristic of electronic property variations. In the case of MoO_2, the crystal structure can vary slightly from tetragonal-rutile-type (a = 5.6 Å, c = 4.9 Å), while MoO_3 is orthorhombic (a = 3.96 Å, b = 13.86 Å, c = 3.70 Å; S.G. Pbnm).

Perovskite (ABO_3) and Spinel (AB_2O_4) Structures

Perovskite and spinel crystal structures are both ideally cubic and have a wide range of engineering applications. The perovskite structure with the two equivalent cubic (S.G. $Pm\bar{3}m$) forms shown in Fig. 4 has the capacity to incorporate essentially all of the elements in the periodic table (with the exception of the noble gas elements and beryllium and phosphorus) (Mitchell 2002). Obviously, this results in perovskite-structured materials exhibiting a wide range of dielectric, piezoelectric, optical,

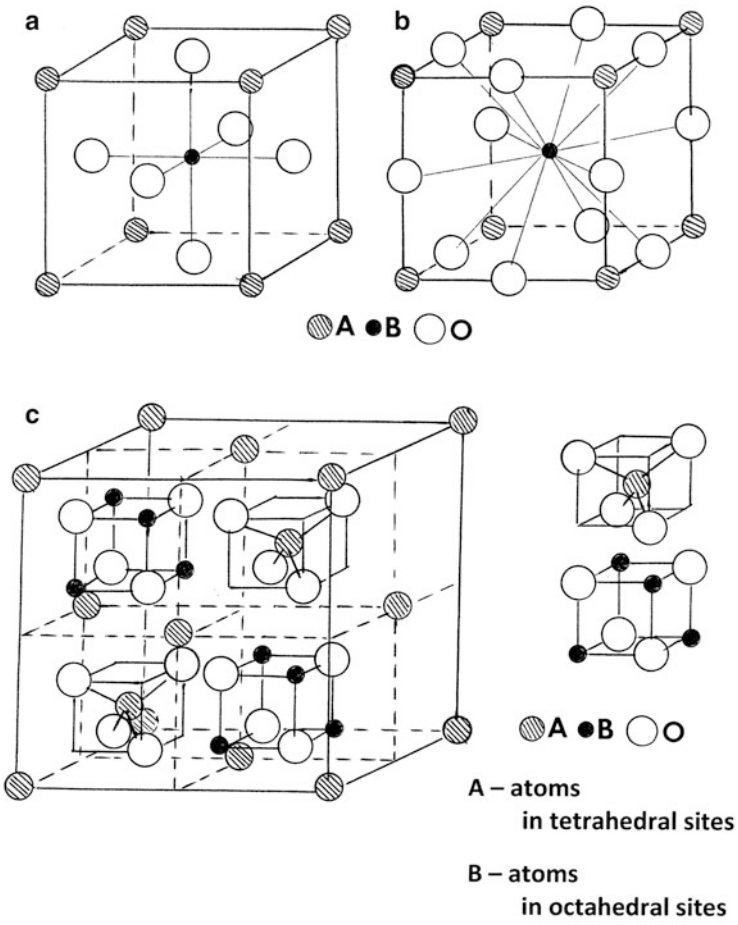


Fig. 4 Perovskite (equivalent) structure (ABO_3) (**a**) and (**b**) and spinel structure (AB_2O_4) (**c**). The unit cell is divided into eight-cube subunits. The four front subunits shown contain $A_4B_8O_{16}$. The complete unit cell therefore is represented by $8(AB_2O_4)$. Note that while oxygen is shown larger than cations, it is actually smaller than most (see ▶ Fig. 7 in chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements")

ferroelectric, magnetoresistive, and superconducting properties. Correspondingly, cubic spinel structures (S.G. Fd3m), often called cubic ferrites, are used in soft magnetic applications such as transformer cores, inductors, memory devices, and high-frequency applications such as microwave systems. Some spinels also exhibit low-temperature superconductivity. The iron oxidation states $FeO \rightarrow Fe_2O_3$ (hematite) $\rightarrow Fe_3O_4$ (magnetite); $Fe^{2+}Fe_2{}^{3+}O_4{}^{-2}$ illustrates the natural occurrence of cubic ferrite and the structural variation of magnetic properties. However, unlike perovskite, the cubic spinel structure is considerably more complex as illustrated in Fig. 4. The spinel structure consists of eight fcc unit cells of oxygen atoms with the

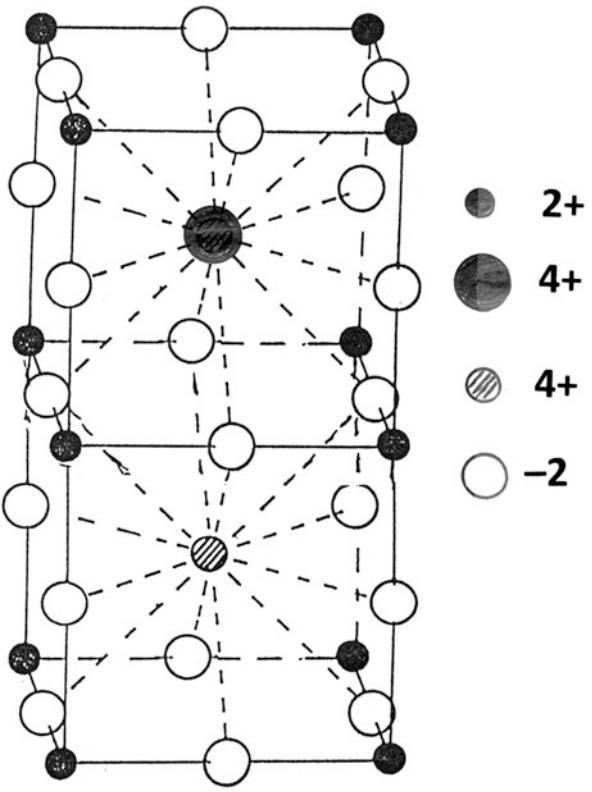


Fig. 5 Double perovskite (tetragonal) unit cell containing different cation species

AB_2 cations located in the tetrahedral and octahedral interstitial positions in this structure. As illustrated for $Fe^{2+}Fe_2{}^{3+}O_4{}^{-2}$, the trivalent ($Fe^{3+}$) ions (or trivalent B ions) located in the octahedral sites have their magnetic moments aligned with an external magnetic field, while the (Fe^{2+}) ions (or divalent A ions) located in the tetrahedral sites have their moments aligned opposite to an external field.

The original perovskite mineral, $CaTiO_3$, was succeeded by $BaTiO_3$ which was a slightly orthorhombic structure, allowing for the Ti ion to create polarization of the unit cells and inducing a significant piezoelectric effect. Over the years, the physical properties of perovskite structures have been shown to depend on stoichiometric variations of A and B ions and charge balance features associated with both cationic and anionic vacancies. In addition, double perovskites have been explored in connection with magnetic properties; the double perovskite Sr_2YRuO_6 is an antiferromagnetic insulator, and similar properties are observed in $SrLaMnRuO_6$ and related compositions (Fig. 5). Tetragonal variants (double perovskites) with ferroelectric properties include $Sr_{1-x}Ba_xNb_2O_6$. A simpler and equally novel perovskite system involves WO_3 (fcc) with a neutral metal, such as sodium, in the cell center, creating the perovskite structure shown in Fig. 4a. This structure was first recognized by F. Wohler in 1924 and became known as sodium tungsten bronze:

Na_xWO_3. For $x = 0.9$, neighboring unit cells contain a vacancy, and the material appears golden yellow. The color changes to orange for $x = 0.8$, red for $x = 0.6$, purple for $x = 0.5$, and deep blue for $x = 0.4$. The structure has been observed to be superconducting at $x = 0.3$, but with a very low $T_c = 0.6$ K. Like other charge-balance vacancy structures, the color change is associated with decreasing conductivity and structural deviation from cubic perovskite. For $x = 1$ ($Na^0W^{6+}O_3{}^{-2}$), the material has a conductivity roughly one-third that for metallic sodium; decreasing for decreasing sodium concentrations. These structures can be ordered: ordered sodium and ordered vacancies. For $x = 0.5$, the ordering produces interpenetrating Na and vacancy fcc superlattices: neutral sodium in every other perovskite unit cell (Fig. 4a). It might be recalled that for F-center formation in NaCl (▶ Fig. 10b in chapter "Electromagnetic Color and Color in Materials"), the electron occupying the Cl (anion) site can combine with Na^+ to form neutral sodium ($Na^+ + \bar{e} = Na^0$) and the sodium atoms can agglomerate to form colloidal sodium, or sodium metal nanoclusters. This can be envisioned for sodium-tungsten bronze where Na^+ can occupy one unit cell and an electron the other. Alternatively, we can have $[Na^+_X W^{5+}_x W^{6+}_{(1-x)}]O^{-2}_3$, where above a threshold of $x > 0.20$, the structure becomes conducting. For $x = 1$, the $(Na^+W^{5+}) \rightarrow (Na^0W^6)$ by inter-valence transfer. These structures have been variously studied for nearly a century.

In retrospect, perovskite or perovskite structure is recognized as a novel material host since it can be variously doped and configured as a double unit cell or even a triple unit cell with tetragonal structure, $c \cong 3a$, or orthorhombic structure depending on oxygen vacancies accommodating charge balance necessary for mixed-valence cation doping. This is especially apparent in the double perovskite structure shown in Fig. 5 whose cation valences are 2+ and 4+. For a substitution of a 2+ or 3+ cation into either half-cell center would require correspondingly a charge reduction from +12 to +6 or +11, respectively. This would require a reduction in the double perovskite cell, or 0.5 and 0.25 oxygen vacancies in either half cell, with characteristic property changes for these oxygen-defect perovskite structures.

References

Mitchell RH (2002) Perovskites: modern and ancient. Almaz Press, Thunder Bay

Structures and Properties of Oxide Superconductors

Contents

Abstract

Using chapter "▶ Electrovalent Crystal Structures and Chemistry," and especially the crystal chemistry of metal oxides as a platform, this chapter will examine basic structures and properties of oxide superconductors, particularly the so-called high-temperature (HT) superconductors. The two conditions for true superconductors, zero resistance (or resistivity), and a demonstrated Meissner effect (levitation of a magnet) are described, the latter as the exclusion of magnetic flux in the superconducting state. However, this chapter begins with a brief overview of more conventional metal or alloy superconductors in order to establish a relationship between these more conventional superconductors in contrast to oxide superconductors. It is pointed out that since the popular theory of superconductivity, the BCS theory, does not elucidate specific mechanisms which provide any predictability or avenues for guided exploration, future innovations may hinge on novel design strategies to improve existing systems.

Introduction

Superconductivity was discovered around 1908 by Kamerlingh Onnes only because he had discovered how to liquefy helium (4.2 K). He simply exercised his curiosity regarding the behavior of metals in general and initially mercury in particular when

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_14

immersed in liquid helium. He initially found that the electrical resistance of mercury became effectively zero in liquid helium and later found the same phenomenon for other metals such as lead and tin.

From about 1915 through the mid-1960s, metal and alloy superconducting materials were developed and explored. Niobium, superconducting at 7 K, was investigated between 1930 and 1945. Higher-temperature Nb_3Sn (superconducting at ~18 K) and Nb_3Ge (superconducting at ~23 K) were developed between 1955 and 1965. These could be fabricated as wire systems. Oxide superconductors exhibiting some similarities to Nb, referred to as bronzes, $AxWO_3$, $AxMoO_3$, and $AxReO_3$, for example, were explored in the 1960s, with superconductivities as superconducting transition temperatures ranging from ~4 to 6 K. $LiTi_2O_4$ and Ba (Pb, Bi)O_3 were shown to superconduct at ~13 K around 1975, where the A_3B metal alloy systems continued to exhibit superconductivity around 23 K.

In 1986, essentially by accident, IBM-Zurich scientists, G. Bednorz and A. Müller, discovered that $(La, Sr)_2CuO_4$ exhibited superconductivity at ~35 K, and only a year later $YBa_2Cu_3O_7$ was observed by M. K. Wu and C. W. Chu to superconduct at ~95 K, which allowed the superconducting state to be achieved using liquid nitrogen (~78 K). The interesting feature of $YBa_2Cu_3O_7$ is that its unit all structure was basically a stack of three perovskite unit cells, but the stoichiometry required systematic oxygen (anion) vacancies to charge balance the structure, consistent with the crystal chemistry principles demonstrated in chapter "▶ Electrovalent Crystal Structures and Chemistry." In addition, the valence of Cu was observed for the first time to be +3, at variance with +1 and +2 for the prominent copper oxide, $Cu_2^{+1}O^{-2}$ and $Cu^{+2}O^{-2}$.

Other copper oxide superconductors were discovered in the 1980s, with superconducting transition temperatures of around 130 K. But since then, very little progress has been made. Oxide superconductors represent very intriguing materials systems, where there remains much that is unknown, even more than 25 years after the discovery of high-temperature superconducting oxides. In the sections to follow, some of the more interesting oxide superconductor structures and properties are examined as examples of these intriguing materials systems.

Superconductivity Fundamentals in Perspective

In the mid-1908, Kamerlingh Onnes, a Dutch scientist, succeeded in liquefying helium by sequentially liquefying air (~78 K) then hydrogen (~20 K) and then helium (4.2 K). This allowed him to explore very low temperature conduction in metals, and in 1911, in experiments on mercury cooled to ~4.2 K using liquefied helium, he observed a rapid loss of resistance which became essentially zero or immeasurable. He referred to this phenomenon as "supraconductivity" and later "superconductivity", also observing that tin and lead displayed zero resistance, or superconductivity, at 3.7 and 7.2 K, respectively. In 1913, Onnes received the Nobel Prize in physics, not for superconductivity but for the liquefaction of, and experiments with, helium. Continued superconductivity experiments with other metals

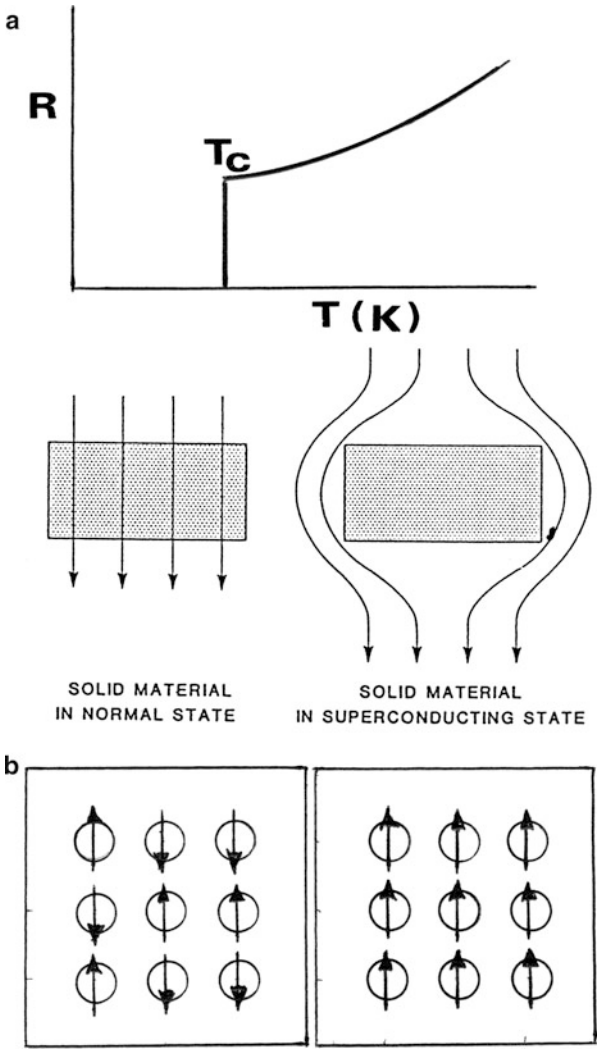


Fig. 1 Conditions for superconductivity. (**a**) Rapid transition to zero resistance at a critical temperature, T_c. (**b**) Meissner effect. Diamagnetic expulsion of magnetic flux lines from the volume of a solid material in the superconducting state

demonstrated a critical temperature (T_c) of 9.2 K for niobium in 1930 (Fig. 1a). In 1933, W. Meissner and R. Ochsenfeld discovered that when a sphere of lead was cooled below its transition temperature (or critical temperature; $T_c = 7.2$ K) in a magnetic field, it excluded the magnetic flux, allowing the material in the superconducting state to levitate in the field. Consequently, this "Meissner effect" became a property unique to the superconducting state (Fig. 1b and c). Superconductors were found to exhibit zero resistance at and below T_c and the Meissner effect (Fig. 1). Superconducting elements which exhibited the Meissner effect illustrated in Fig. 1b and c became known as type I superconductors, while later, alloy superconductors, which did not completely exclude magnetic flux lines, became known as type II superconductors (Wilson 1983).

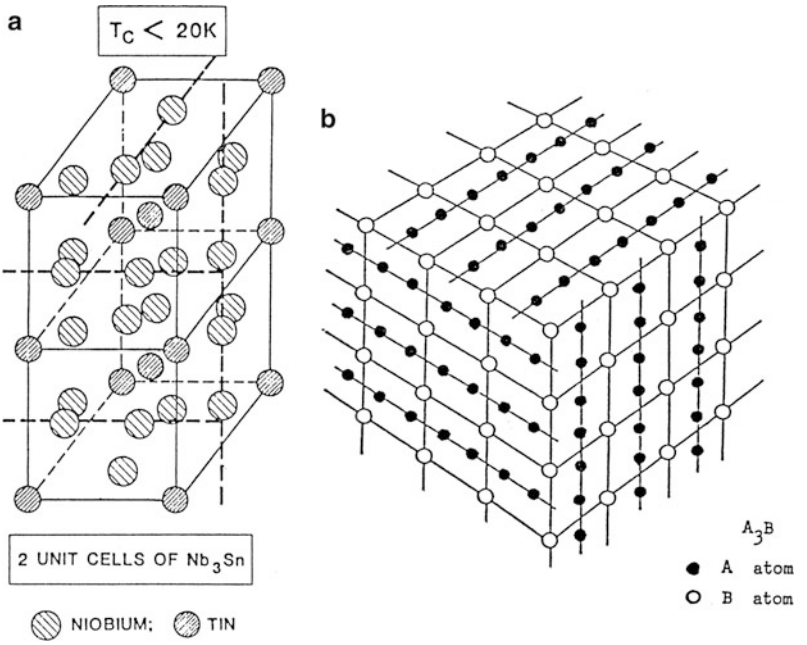


Fig. 2 Cubic (A15) crystal structure for A_3B, type II superconducting alloys, (**a**) 2 unit cells of Nb_3Sn. (**b**) multiunit-cell arrangement for A_3B A15 cubic structure illustrating Nb chains

The exclusion of magnetic flux in connection with superconductivity was in fact observed by K. Onnes when he discovered that applying a sufficiently strong magnetic field (eventually described as a critical field, H_c) to a superconducting element restored the resistance to normal, effectively quenching the superconducting state. Much later observations also illustrated a similar quenching of superconductivity with high current having a superconducting wire, referred to as a critical current or current density, J_c:

$$J_c = \frac{2}{3}\left(1 - \frac{T}{T_c}\right)^{\frac{3}{2}}\left[\frac{10H_c}{4\pi\wedge}\right], \tag{1}$$

where $\wedge$ represents the penetration depth of the magnetic flux into the superconductors.

As shown in Fig. 1, the normal state of a type I superconductor can be represented by a random arrangement of atomic dipoles, while in the superconducting state the dipoles are oriented in the same direction creating a state of perfect *diamagnetism* and expelling magnetic flux. For single elements, including Nb, the critical current and magnetic field were observed to be so low that practical uses of these type I superconductors were essentially excluded.

In the 1950s, however, superconducting A_3B alloys having a cubic (A15) crystal structure shown in Fig. 2 were discovered to have T_c ranging from 10 to 20 K and correspondingly high critical current densities (J_c) and magnetic field intensities (H_c).

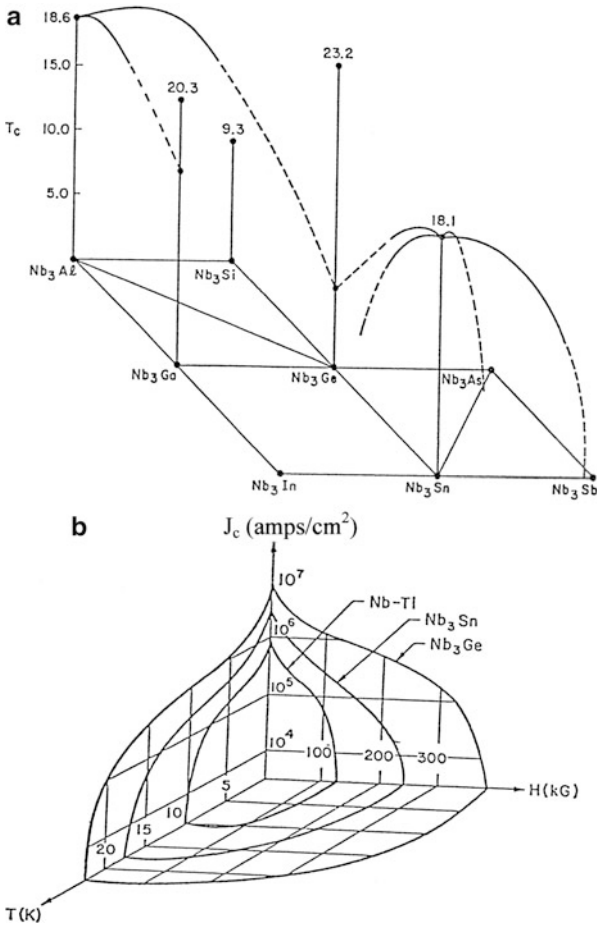


Fig. 3 Niobium-based, binary A15, type II superconducting compounds. (**a**) Solid solution T_c diagram. (**b**) Parameter space for several practical (commercial) type II superconductors

This was particularly true for Nb-based alloys illustrated in Fig. 3, which include Nb_3Ge with $T_c = 23.2$ K, developed in the early 1970s. These alloys (Fig. 3) could be commercially processed to manufacture wire, evolving as arrays of superconducting filaments or overlapping filament fragments in a copper matrix which could allow for excursions from the superconducting state and absorb the associated and fluctuating normal-state heating. Figure 4 illustrates (schematically) the superconducting wire composites which are currently in use in the fabrication of large magnets for MRI systems and related high magnetic field applications.

It can be observed in Fig. 3b that from a practical, operating perspective, optimal conditions for the A_3B superconducting alloys are those within the corresponding enclosed parameter space volume (H_c-J_c-T_c), for Nb_3Ge. In addition, for these type II superconductors, the magnetic field first penetrates the superconductor at low magnetic fields, referred to as the lower critical field, H_{cl}, which does not destroy the superconductivity. However, this magnetic flux line penetration creates normal

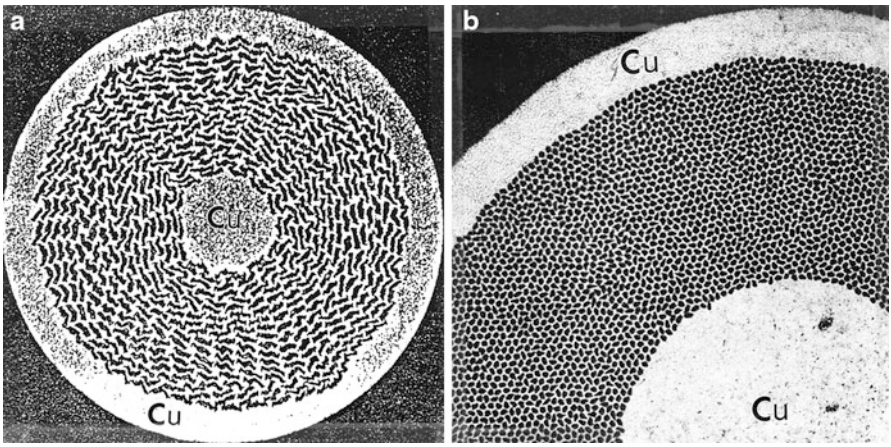


Fig. 4 Cross section of an 8 mm diameter composite fiber, type II superconducting Nb-Ti wire with copper cladding. The copper-to-superconductor ratio is 1.9:1. The Nb-Ti filaments are rolled into the copper to develop what is referred to as a "jelly roll" structure by Teledyne Wah Chang, Albany, Oregon; a major manufacturer of superconducting wire (**a**). (**b**) shows Nb-47 wt. % Ti/Cu composite (1:1.5), 0.8 mm diameter wire. The superconducting filament diameter is 6 mm (courtesy of IGC Advanced Superconductors, Inc)

metal regions in the superconductor which are called vortexes or *fluxoids*. These can be thought of as quantized flux lines. Pinning forces (or clusters) in the superconductor such as grain boundaries, particles, or dislocation substructures keep the fluxoid immobile up to some upper critical field, H_{c2}, where they move, and superconductivity is destroyed. Figure 5a and b illustrates these phenomena. Figure 5b in particular illustrates the increase in the flux pinning force with decreasing grain size, D. Figure 5c and d illustrates the fluxoid partitioning for type II superconductors. It can be observed in Fig. 5c and d that if the fluxoids can be represented by magnetic field loops in a spatial confinement or partitioned microstructure characteristic of smaller grain sizes or increasing density of dislocation substructures, the current in a unit area of superconductor will increase, or the current density, J, will increase. Conversely, defects will enhance superconduction, which is the opposite for normal-state conduction, where the insertion of defects into the material will decrease the electron mean free path, decreasing the conductivity (▶ Eq. 1; chapter "Electromagnetic Color and Color in Materials").

As illustrated in Fig. 1, the superconducting state induces perfect diamagnetism where atomic dipole moment alignment expels (or repels) magnetic flux lines. In addition, early research on type II superconductors indicated that superconductivity favored low symmetry (cubic) structures and chain-like Nb atom alignment as shown in Fig. 2. But this does not in itself provide a mechanism or theory for superconductivity.

The popular theory of superconductivity was posed by John Bardeen, a corecipient of the Nobel Prize in physics in 1956 (shared with W. Shockley and J. Brattain) for the

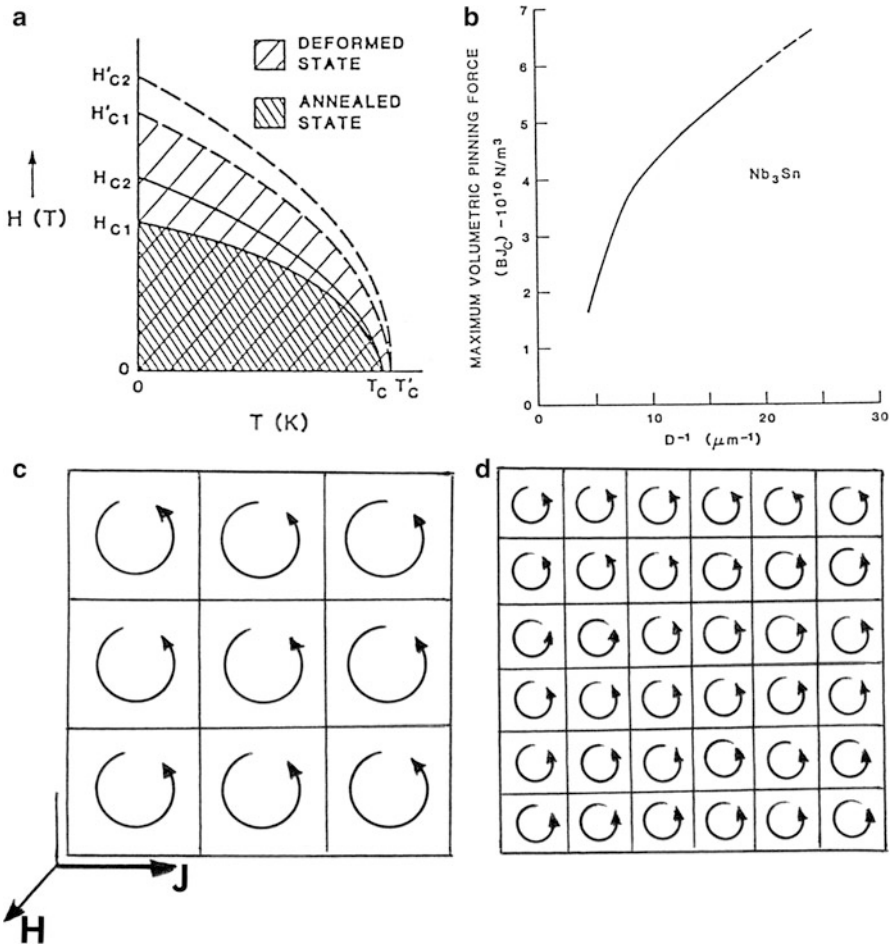


Fig. 5 Influence of deformation or increasing defect density on upper and lower critical field (**a**) and decreasing grain size (D) on flux pinning (**b**). (**c**) and (**d**) show the concept of increasing current density for decreasing microstructure, such as grain size (D)

invention of the transistor at Bell Laboratories more than 5 years before, and J. R. Schrieffer (his graduate student) and L. Cooper (his post doc) at The University of Illinois, around 1957. This theory became shown as the *BCS theory* of superconductivity (Bardeen et al. 1957) and Bardeen, Cooper, and Schrieffer shared the Nobel Prize in physics in 1972 for this work, making Bardeen the first recipient of two Nobel Prizes in the same field (physics).

A principal feature of the BCS theory is the formation of Cooper pairs or paired electrons in the superconducting state. These paired electrons have quantized energy states which are unable to absorb the energy of the normal scattering mechanisms experienced by conduction electrons shown schematically

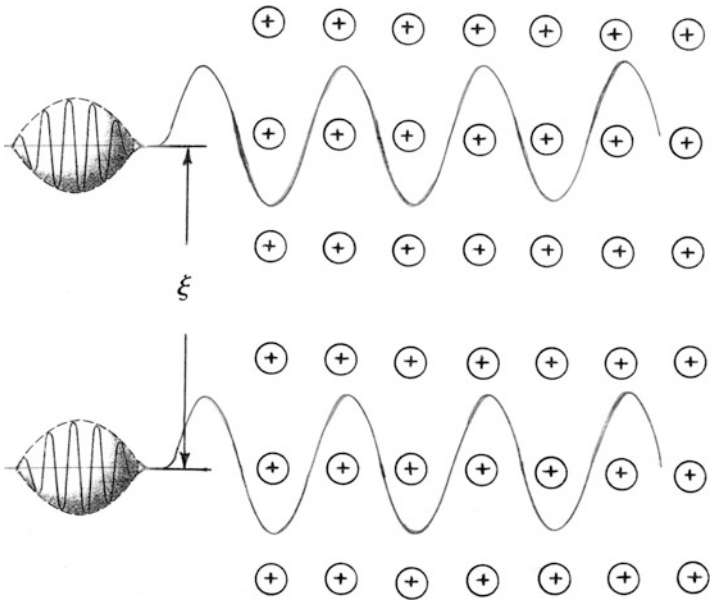


Fig. 6 Simple quantum mechanical model for electron pair-wave propagation in a crystal lattice. The coherence length is indicated by ξ

in ▶ Fig. 5c in chapter "Electromagnetic Color and Color in Materials." These paired electrons can effectively move freely through the superconducting solid, and in effect the mean free path (▶ Eq. 1 in chapter "Electromagnetic Color and Color in Materials") becomes infinite, and the conductivity correspondingly becomes infinite. In addition, these electron pairs were held together by an attractive interaction through coupling to vibrations (or phonons) of the crystal lattice. So long as the paired electrons remained "together," superconductivity was maintained. The pairing distance was demonstrated to be characterized by a *coherence length*. For coherence lengths larger than normal-state discontinuities in a regime, superconductivity would be maintained. However, for discontinuous length scales greater than ξ superconductivity ceased. In a practical sense, the superconducting fibers in the copper matrix shown in Fig. 4 must not be separated more than ξ, the shortest distance over which superconductivity can be maintained.

In some respects, this concept of electron pairing might also be viewed as a quantum mechanical transformation, where in the normal conduction state, electrons behave as "particles" moving in an electric field, while in the superconducting state, the paired electrons behave as "waves" interacting with the lattice atoms. This is illustrated phenomenologically in the cartoon of Fig. 6. The phonon interaction shown in Fig. 6 is a characteristic of the Nb chains illustrated in Fig. 2b.

Of course in the context of some kind of energy difference or superconducting energy bond gap differentiating the transition from the normal state to the superconducting state (at the top of the Fermi level (E_F))

$$E_g = 4E_o \exp(-1/\lambda), \tag{2}$$

where E_O is the Debye energy $k_B\theta_D$ characteristic of the lattice vibrations or displacement of the atoms (+) in chapter "▶ Electromagnetic Color and Color in Materials" Fig. 46, k_B is the Boltzmann constant, and θ_D is the Debye temperature. λ in Eq. 2 is called the electron–phonon coupling constant which tends to increase with increasing T_C:

$$T_C = 1.134\,\theta_D \exp(-1/\lambda). \tag{3}$$

Dividing Eq. 2 by Eq. 3 provides the universal dimensionless ratio:

$$E_g/k_B T_C = 3.53. \tag{4}$$

The critical temperature can also be represented by

$$T_C = 1.14\,\theta_D \exp[-1\,(\lambda - \mu^*)], \tag{5}$$

where

$$\mu^* = \mu/[1 + \mu \ln(E_F/E_D)] \tag{6}$$

is referred to as the Coulomb interaction pseudopotential and μ is the dimensionless Coulomb constant which is the product of the screened Coulomb potential $<V_C>$ and the density of states at the Fermi level, $N(E_F)$. The BCS theory also predicts a discontinuity in the electronic contribution to the specific heat at the transition temperature given by

$$\left(\frac{C_s - C_n}{C_n}\right)_{T_c} = 1.43 \tag{7}$$

and

$$T_c \cong 1/\sqrt{M} \tag{8}$$

where in Eq. 7 the subscripts s and n denote the superconducting and normal states, respectively, and $C_n \propto T_c$. M in Eq. 8 is the average isotopic mass.

Appropriate substitutions in Eqs. 3, 4, 5, and 8 indicate a maximum $T_c \sim 30$ K, which for type II superconducting alloys has prevailed. On the other hand, there are no aspects of the BCS theory which are predictive in the sense that some avenue of exploration is evident (Etourneau 1992; Parks 1969).

Superconductivity in Oxide Systems

As noted previously, tungsten bronzes exhibited superconductivity in the 1960s along with simple oxides such as TiO and NbO with T_cs below 1 K. In early 1973, $LiTi_2O_4$ spinel (▶ Fig. 4c in chapter "Electrovalent Crystal Structures and Chemistry") was observed to exhibit superconductivity at ~14 K followed by chalcogenide spinels

such as $CuRhS_4$ and $CuRh_2Se_4$ at lower T_cs, while around the same time Ba (Pb, Bi) O_3 (perovskite) was found to have $T_c \cong 13$ K. In the 1980s, other chalcogenide superconductors such as $PbMo_6S_8$ (referred to as Chevrel phases) were observed to have superconducting transition temperatures as high as 14 K.

The barium lead bismuth oxide system, made from the perovskite solid solution of insulating $BaBiO_3$ in the metallic $BaPbO_3$, took on a wide range of interest in the 1980s. This was due partly because of the observation that superconductivity appeared at the boundary between the metallic and insulating phases at T_c ~12 K, for a Bi concentration of $x = 0.25$ in the Ba $(Pb_{1-x}Bi_x)$ O_3 mixed perovskite solid solution. Interest also included ternary cuprates like $La_4BaCu_5O_{13}$ and other La-Ba-Cu-O ceramic oxide systems. In 1986, G. Bednorz and A. Müller reported superconductivity in this system at 28 K, 5 K higher than the Nb_3Ge type II alloy superconductor (Fig. 3a) (Bednorz and Müller 1986). Bednorz and Müller shared the 1988 Nobel Prize in physics for their discovery. This critical temperature was quickly raised to 38 K in the tetragonal $La_{1.8}Sr_{0.2}CuO_4$ structure. Soon after these discoveries, M. K. Wu and C. W. Chu demonstrated superconductivity in $YBa_2Cu_3O_7$, a triple perovskite structure, at ~90 K (Chu et al. 1987). This extraordinary elevation in T_c from ~23 K in Nb_3Ge to ~90 K in $YBa_2Cu_3O_7$ within a period of several months at the end of 1986 and the beginning of 1987 was unprecedented and certainly not predicted by the BCS theory of superconductivity where T_c was shown not to exceed ~30 K within a year of the discovery of the triple perovskite, copper-oxide superconductor, 123-$YBa_2Cu_3O_7$, superconductivity was discovered in another exotic Bi/Sr/Ca/Cu-O tetragonal system at 100 K (Maeda et al. 1988), and a Tl/Ba/Ca/Cu-O tetragonal system at ~122 K (Sheng and Hermann 1988).

Figure 7a illustrates the superconducting $YBa_2Cu_3O_7$ orthorhombic structure ($a = 3.82$ Å, $b = 3.89$ Å, $c = 11.68$ Å) composed by a stack of 3 oxygen-defect perovskite cells, ideally represented by the $YBa_2Cu_3O_9$ structure (Fig. 7b), which is nonexistent because it cannot be charge balanced. The notable feature in the $Y^{3+}Ba_2^{2+}$ $(Cu^{+}Cu_2^{3+})$ O_7^{-2} structure is the mixed Cu valence and in particular the observation of Cu^{3+} for the first time. For the tetragonal $YBa_2Cu_3O_6$ structure shown in Fig. 7, it can be charge balanced with $Y^{3+}Ba_2^{2+}Cu^{+}Cu_2^{2+}O_6^{-2}$, also a mixed Cu valence, but without Cu^{3+}. Figure 7d and e shows the superconducting transition for $YBa_2Cu_3O_7$, a semiconducting behavior for $YBa_2Cu_3O_6$, and a metallic conductor behavior for $YBa_2Cu_3O_8$ (structure not shown). It is also notable that, as shown in Fig. 7a, the $YBa_2Cu_3O_7$ superconductor is an oxygen vacancy (defect) structure.

Figure 8a illustrates the Cu-O plane structure changes for stoichiometric variations between $YBa_2Cu_3O_6$: $YBa_2Cu_3O_x$($x = 6$–7), while Fig. 8b illustrates the measured resistance–temperature (R-T) superconductivity profiles for $x = 7$ to $x = 6.3$. It can be noticed that the filled Cu-O planes (or sheets) for the O_7 structure in Fig. 8a are depleted of oxygen for the non-superconducting O_6 structure, but up to $O_{6.25}$ in Fig. 8a, Cu-O chains exist. These Cu-O chains and sheets have often been touted as functioning like the Nb chains in $Nb_3(B)$ type II superconducting alloys (Fig. 2b), A_3B compounds.

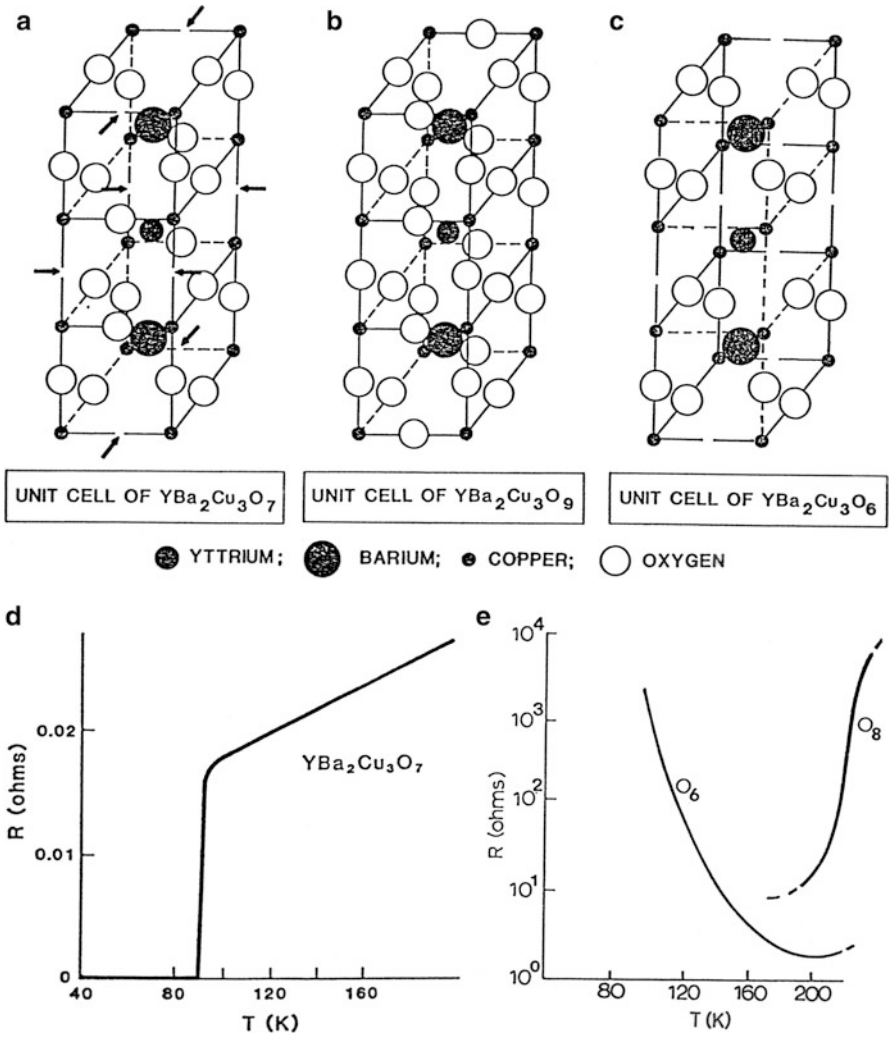


Fig. 7 Copper oxide structures based on three stacked perovskite cells. (**a**) Superconducting $YBa_2Cu_3O_7$ oxygen-defect structure → (orthorhombic unit cell), (**b**) ideal (nonexistent) $YBa_2Cu_3O_9$: $3ABO_3$ tetragonal structure, (**c**) $YBa_2Cu_3O_6$ tetragonal non-superconducting structure, (**d**) R-T plot for superconducting $YBa_2Cu_3O_7$ in (**a**). (**e**) R-T plots for O_6 and O_8 variances of $YBa_2Cu_3O_x$

While the superconducting $YBa_2Cu_3O_7$ structure (Fig. 7a) is orthorhombic, other structures, including the non-superconducting structures, are tetragonal. The more complex Bi-Sr-Ca-O and Tl-Ba-Ca-Cu-O superconducting systems are tetragonal and with increasing c-axis dimension with increasing oxygen content. Correspondingly, in these systems, the critical temperature is also observed to increase with increasing oxygen content. Table 1 illustrates these features for some representative systems,

Table 1 Compositions and transition temperatures for tetragonal oxide superconductors

Superconducting compound	T_c(K)
$Bi_2Sr_2CuO_6$	12
$Bi_2Sr_2CaCu_2O_8$	90
$Bi_2Sr_2Ca_2Cu_3O_{10}$	110
$T\ell_2Ba_2CuO_6$	90
$T\ell_2Ba_2CaCu_2O_8$	110
$T\ell_2Ba_2Ca_2Cu_3O_{10}$	125
$T\ell Ba_2CaCu_2O_7$	80
$T\ell Ba_2Ca_2Cu_3O_9$	110
$T\ell Ba_2Ca_2Cu_4O_{11}$	122
$(T\ell_{0.5}Pb_{0.5})Sr_2CaCu_2O_7$	90
$(T\ell_{0.5}Pb_{0.5})Sr_2Ca_2Cu_3O_9$	122

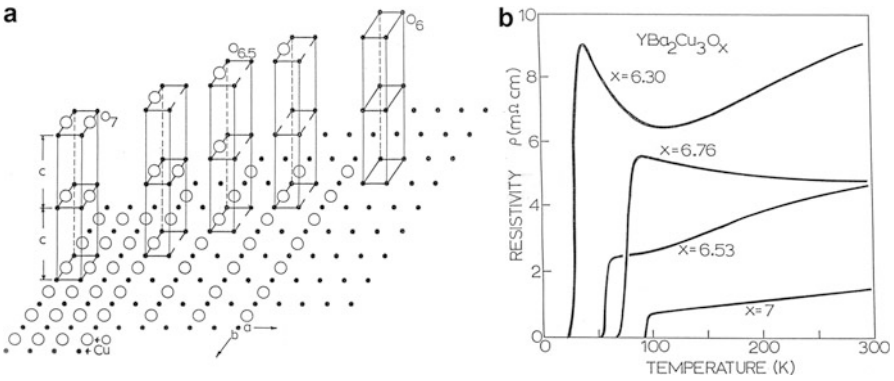


Fig. 8 (**a**) Oxygen-defect $YBa_2Cu_3O_x$ double, partial cell arrays (*c* is the unit cell dimension) showing intercell Cu-O sheet and chain structures for x = 7 to x = 6. (**b**) measured resistivity versus temperature response for O_x structures showing the superconducting transitions

while Figs. 9 and 10 provide a few examples of the complexity of the corresponding tetragonal crystal structures for the high-T_c superconducting Bi- and Tℓ-based copper oxides (Park and Snyder 1995).

In 1994, mercury was substituted for Tℓ in $TlBa_2Cu_3O_9$ forming Hg $Ba_2Ca_2Cu_3O_9$ which was observed to be superconducting at 134 K. This critical temperature rose to 164 K under pressure and represents the highest reproducible T_c recorded to date (Hunter et al. 1994). Pressure was used in the development of metal alloys and other compounds in attempts to force lower symmetry and more stable (cubic) crystal structures such as the A15 structure (Fig. 2).

Although the BCS theory of superconductivity does not apply in any significant way to high-T_c superconductors or high-T_c superconductivity, there have been no significant theoretical contributions pointing the way toward new or higher-T_c cuprate superconductors or little insight into the prospects for finding new or more exotic superconductors (Komimura et al. 2010). Despite these limitations,

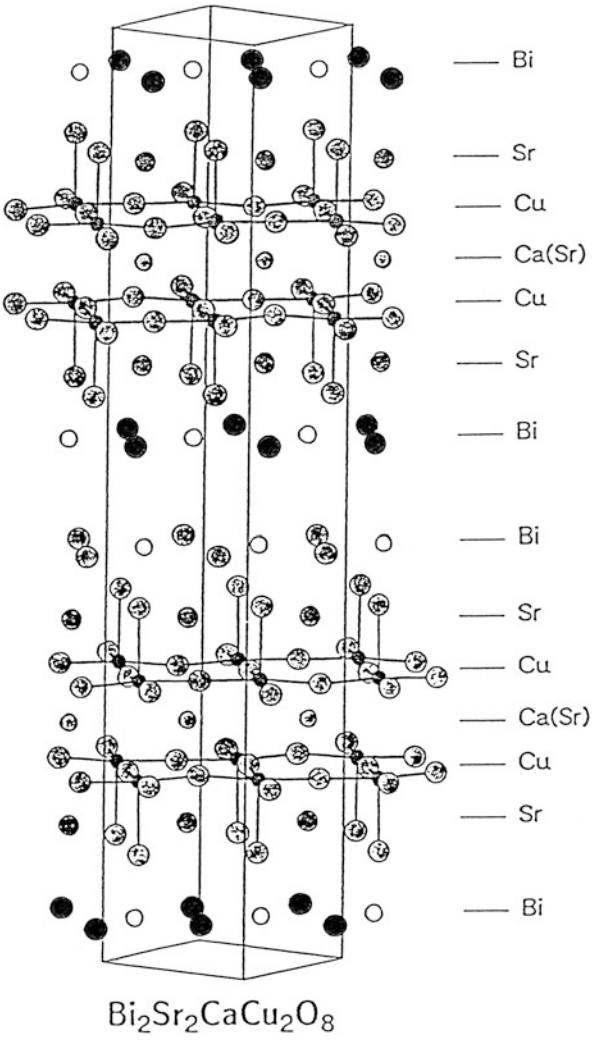


Fig. 9 Tetragonal $BiSr_2CaCu_2O_8$ structure. Note Bi-O layers

superconductivity continues to have an impact on our lives, albeit limited and specialized, and involving the A15 alloys, especially Nb-Ti composite wire forms illustrated in Fig. 4. Despite only small improvements in T_c or even current carrying capacity (J_c) (Fig. 3b), helium production, storage, and transfer have made a significant difference.

The high-T_c tetragonal superconductors (Fig. 10) have particular challenges because of the enormous c-axis dimensions and correspondingly small coherence lengths, which in general are an order of magnitude smaller than the A15 alloys. In addition to the very small coherence lengths, they are non-isotropic, with order-of-magnitude differences between the c-axis direction and the a- and b-axis directions

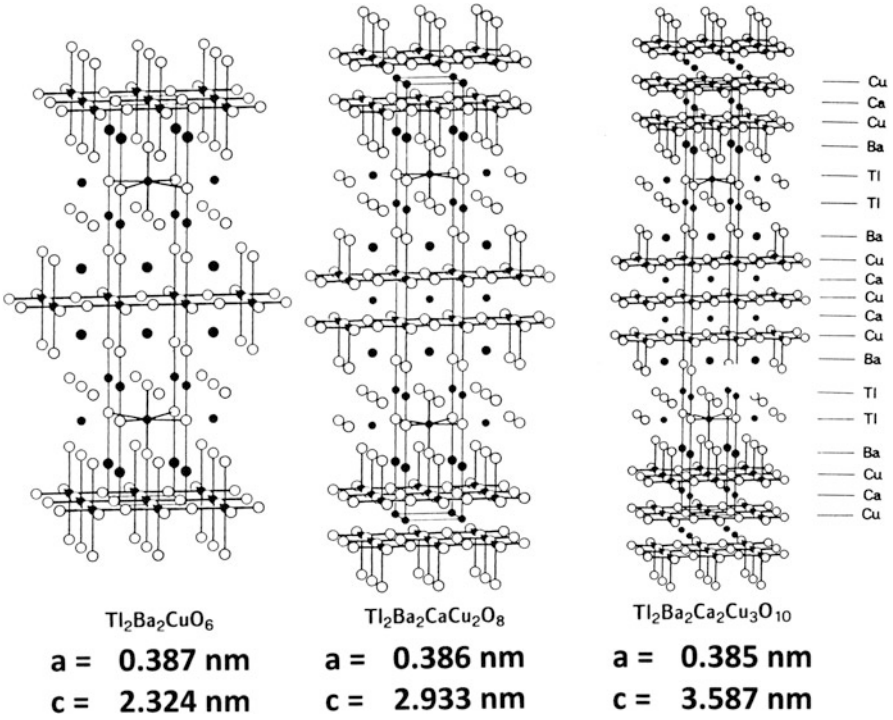


Fig. 10 Examples of Tl-Ba-(Ca)-Cu-O superconductor structures: $Tl_2Ba_2Ca_{n-1}Cu_nO_{2n+4}$; n = 1, 2, 3. Note open circles are oxygen

(lattice parameters). Consequently, polycrystalline high-T_c oxide superconductors require large oriented or textured configurations, where even grain boundaries can interrupt superconductivity since the effective grain boundary width can exceed the coherence length. Table 2 compares coherence lengths for a number of superconductors along with the corresponding critical temperatures, T_c.

In addition to coherence length issues, high-T_c oxide superconductors are also difficult to fabricate as a consequence of their reactivity, and they are brittle in any form. Consequently, they are difficult to manufacture in convenient wire form and must be fabricated in specific, conformal geometries such as storage rings or other simple shapes, or special sheath or tape form with limited flexure, although some high-T_c superconducting oxide tapes are produced commercially. Like more conventional superconductors, some thin-film applications have been demonstrated. The more important thin-film configurations for superconductors have been Josephson junction arrays illustrated schematically in Fig. 11. J.D. Josephson developed these thin-film superconducting devices in his PhD dissertation research and was awarded a Nobel Prize in physics in 1974. These thin-film devices are composed of superconductors separated by an insulating barrier through which the paired electrons can tunnel to alter the I–V switching

Table 2 Comparison of coherence lengths and superconducting critical temperatures for superconducting materials

Superconducting material	Crystal structure	T_c (K)	Coherence length (ξ) (nm)	
Type I			a	c
Aluminum	Fcc	1.2	1,400	
Tin	Fcc	7.2	80	
Lead	Bcc	9.3	38	
Niobium[a]	Tetragonal	3.7	150	
Type II				
Nb-Ti	Cubic (A15)	13	3	
Nb_3Sn	Cubic (A15)	18	4	
Nb_3Ge	Cubic (A15)	23.2	4	
Oxide (*Type II*)				
$YBa_2Cu_3O_7$	Orthorhombic	90	2	0.4
$Bi_2Sr_2CaCu_2O_8$ (Bi-2212)	Tetragonal	90	1.5	0.1
$Bi_2Sr_2Ca_2Cu_3O_{10}$ (Bi-2223)	Tetragonal	110	1.3	0.1
$T\ell Ba_2Ca_2Cu_4O_{11}$	Tetragonal	122	1.4	0.1
$HgBa_2Ca_2Cu_3O_9$	Tetragonal	134	1.3	0.2

[a]Nb is considered to be a type II superconductor

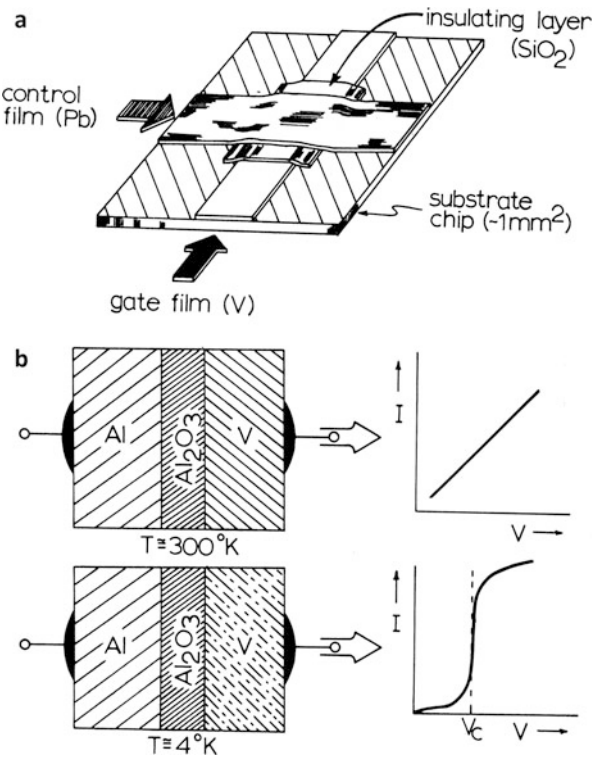


Fig. 11 Superconducting thin-film device phenomena. (**a**) Superconducting thin-film gate device ($Pb/SiO_2/V$: superconductor/insulator/superconductor. Gate current drops to zero when control current exceeds the critical current for the gate film having high normal-state resistance. (**b**) Electron tunneling sandwich as in (**a**) showing normal-state ohmic I–V characteristic and I–V switching in superconducting state

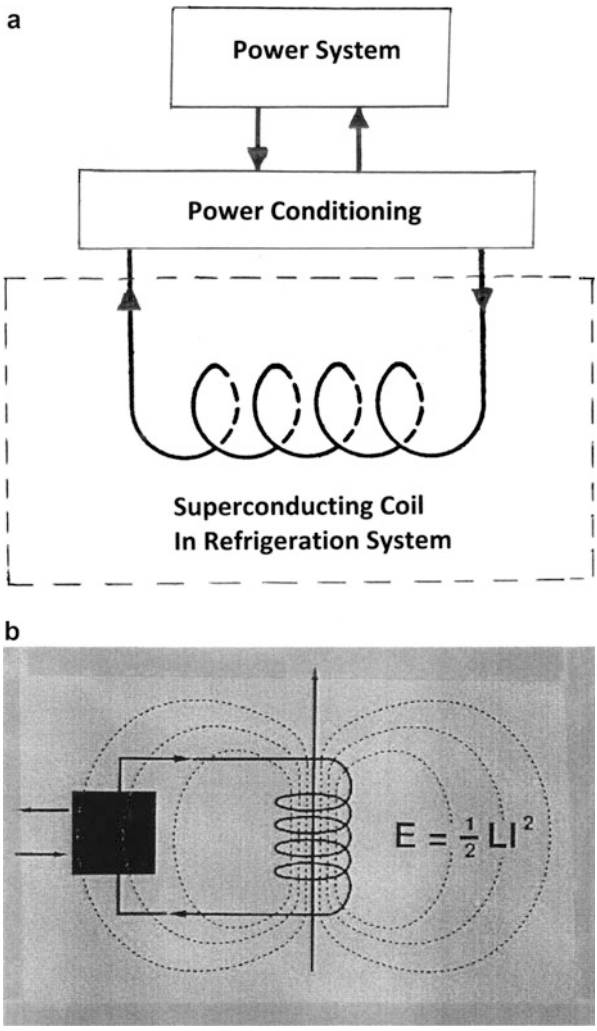


Fig. 12 Schematic for a SMES. (**a**) System schematic. (**b**) Induction coil schematic. The total energy storage, E, is related to the coil inductance, L. L is proportional to the number of turns (N) of conductor squared: $L \alpha N^2$

feature illustrated in Fig. 11b. Applications for these devices include superconducting quantum interference devices (SQUIDS).

Superconducting magnetic energy storage (SMES) is the only known method of bulk storage of electrical energy. Although not widely used, prototypes date back to the late 1960s with demonstration units built in several parts of the world. SMES stores electricity from an electrical grid within the magnetic field of a coil of superconductor. This stored current can be instantaneously discharged back into the electric grid. Figure 12 illustrates this storage concept schematically. Interestingly, and in spite of engineering challenges in using high-T_c oxide superconductors, there have been a half-dozen demonstration SMES high-T_c oxide systems constructed between 2004 and 2012: China, Japan, France, and Korea.

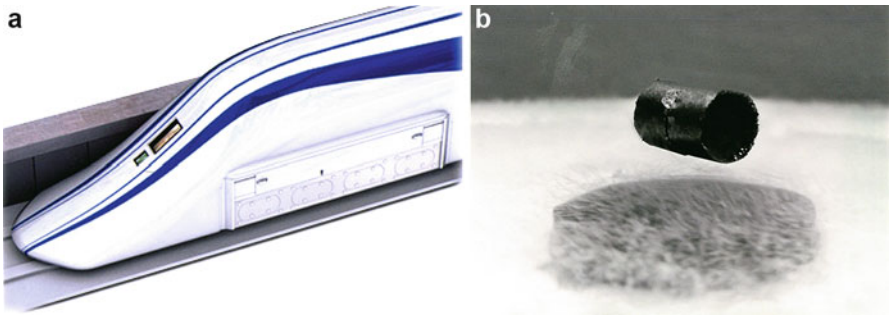


Fig. 13 Superconducting maglev train (**a**). (**b**) shows a magnetic levitating over a $YBa_2Cu_3O_7$ coupon in liquid nitrogen by the Meissner effect shown schematically in Fig. 1b

Two systems used $YBa_2Cu_3O_7$, two used Bi-2212, and two used Bi-2223 (Table 2). These systems ranged from 1 to 2.5 MJ, could handle current densities $\geq 10^6$ amps/cm^2 (Fig. 3b; Ali et al. 2010), and could be cooled with liquid hydrogen (20 K). This provides the prospect for SMES systems in vehicles, especially those which would utilize liquid hydrogen storage for hydrogen-powered vehicles (Morandi et al. 2012).

A more tested use of large superconducting magnetic systems involves levitated trains, often referred to as magnetic levitation or maglev. Systems utilizing conventional electromagnetic levitation systems have been in use in various parts of the world for several decades, but systems utilizing superconducting metal alloys and high-T_c superconductors have also been developed. These systems utilize superconducting magnets attached to the train and oriented toward the rails below, creating a Meissner-induced levitation. A Japanese superconducting maglev prototype reached a speed of 360 mph (580 km/h) in 2003. These systems, illustrated somewhat conceptually in Fig. 13, must maintain a constant distance from train to track of ~1.5 cm and require complex feedback mechanisms to maintain controllable speed and overall system stability.

While it does not seem likely that higher T_c and more tractable superconducting systems will emerge in the near future, progress in engineering efficiencies and innovations in superconductor design may have important applications (Bray 2008). However, in the absence of a more fundamental understanding of superconductivity, progress may be slow unless some serendipitous circumstance presents a new realm of exotic superconductors.

References

Ali MH, Wu B, Dougal RH (2010) An overview of SMES applications in power and energy systems. IEEE Trans Sust Energ 1(1):38–47

Bardeen J, Cooper LH, Schrieffer JR (1957) Microscopic theory of superconductivity. Phyc Rev 106:162–164

Bednorz JG, Müller KA (1986) Possible high T_c superconductivity in the La-Ba-Cu-O system. Z Phys B Con Mat 64:188–190

Bray JW (2008) Superconducting applications: present and future. J Supercond Nov Magn 21 (6):335–341

Chu CW, Hor PH, Meng RC, Gao L, Hwang ZJ (1987) Superconductivity at 52.2K in the lanthanum-barium- copper-oxide system. Science 235:567–569

Etourneau J (1992) Superconducting materials. In: Cheetham K, Day P (eds) Solid state chemistry: compounds. Clarendon, Oxford, UK, pp 60–111

Hunter BA, Jorgensen JD, Wagner JL, Radaelli PG, Hinks DG, Shaked H, Hitterman RL, Von Dreele RB (1994) Pressure-induced structural changes in superconducting $HgBa_2Ca_{(n-1)}Cu_nO_{2n+2+d}$($n = 1, 2, 3$). Physica C Supercond 221:1–8

Komimura H, Ushio H, Matsuno S, Hamada T (2010) Theory of copper oxide superconductors. Springer, Berlin/Heidelberg

Maeda H, Tanaka Y, Fukutomi M, Asano T (1988) A new high-T_C oxide superconductor without a rare earth element. JPN J Appl Phys 27:L209–L210

Morandi A, Trevisani L, Negrini F, Ribani PL, Fabbri M (2012) Feasibility of superconducting magnetic energy storage on board ground vehicles with present state-of-the-art superconductors. IEEE Trans Appl Supercon, 22(2):120–122

Park C, Snyder RL (1995) Structures of high-temperature cuprate superconductors. J Amer Ceramic Soc 28(12):3171–3194

Parks RD (ed) (1969) Superconductivity. Marcel Dekker, New York

Sheng ZZ, Hermann AM (1988) Bulk superconductivity at 120 K in the Tℓ-Ca-Ba-Cu-O system. Nature 332:138–139

Wilson M (1983) Superconducting magnets. Clarendon, Oxford, p 295

Part V

Crystal Imperfections in Engineering Materials

Point Defects

Contents

Abstract

In dealing with materials fundamentals, crystal imperfections or defects represent the most significant aspect of crystal structure or microstructure. Point defects or zero-dimensional imperfections have already been illustrated to determine the properties of extrinsic semiconductors (n-or-p type) (▶ Fig. 8 of chapter "Electromagnetic Color and Color in Materials"), the color of crystals (▶ Fig. 11 of chapter "Electromagnetic Color and Color in Materials"), and the superconductivity in oxide superconductors (▶ Fig. 5 of chapter "Structures and Properties of Oxide Superconductors"). The role of line imperfections (dislocation lines) or one-dimensional defects was implicated in the enhancement of flux or vortex pinning and improved supercurrent density in superconductors, along with the implications of grain size or grain boundaries (planar defects or 2-dimensional defects) (▶ Fig. 5 of chapter "Structures and Properties of Oxide Superconductors"). The formation of point defects and their agglomeration to create voids or 3-dimensional imperfections in the self-irradiation of plutonium and its alloys was also briefly described in Chapter "▶ Structure of Metals and Alloys." This chapter will summarize point defects as crystal imperfections and provide some additional and even unconventional examples of the role they play in determining material properties and performance, especially polycrystalline engineering materials, and with emphasis on metals and alloys.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_43

Introduction

We have already discussed aspects of point defects such as vacancies and interstitials, color centers, and substitutional impurities (or foreign atoms) sitting on lattice sites in a crystal structure (▶ Figs. 7, ▶ 10, and ▶ 11 of chapter "Electromagnetic Color and Color in Materials"), including order and disorder phenomena (▶ Fig. 9 of chapter "Lattice Directions and Planes, and Diffraction by Crystals"). Figure 1 summarizes these point defect phenomena: vacancy and interstitial conventions in Fig. 1a, b and a substitutional impurity in Fig. 1c. Figure 1d illustrates both vacancy diffusion and the substitutional impurity diffusion in concert with vacancy diffusion. The equilibrium ratio of vacancies to atoms is

$$\frac{n_v}{n_o} = e^{(-Q_F/K_B T)} \tag{1}$$

where Q_F is the activation energy for vacancy formation and K_B is the Boltzmann constant. Vacancies migrate in a crystal lattice as atoms jump to vacant lattice sites where the number of such jumps per atom is

$$r_a = (n_v/n_o)\, A e^{-(Q_m/K_B T)} \tag{2}$$

where Q_m is the activation energy for vacancy migration. Equations 1 and 2 both illustrate that vacancy formation and migration are highly temperature (T) dependent.

Figure 1e, f illustrates the formation of vacancies and interstitials in an electrovalent crystal lattice where charge balance must be maintained. Correspondingly, the anion and cation are synergistically involved, and corresponding Schottky defect (molecular vacancy) activation energy is higher than a single vacancy as in Eq. 2. Schottky defect diffusion is also more complicated.

Figure 1g–k illustrates the concept of solute and solvent atoms in neutral metal alloys forming solid solutions discussed previously (▶ Fig. 9 of chapter "Lattice Directions and Planes, and Diffraction by Crystals"). In Fig. 1g solute atoms substitute for solvent (or) atoms or fit into the interstitial sites within the solvent lattice. Figure 1h, j show random and ordered substitution of solute atoms, respectively, while Fig. 1j, k show random interstitial solute atoms and substitutional solute clusters, respectively.

It can be observed in Fig. 1h–k that the presence of vacancies as implicit in Fig. 1d will complicate the point defect structure by allowing for segregation or other related phenomena, including precipitate (or second-phase) formation (zero-dimension defects forming 3D-defects) and corresponding variations for the associated activation energies, implicit in Eqs. 1 and 2.

There are many variations of these vacancy-interstitial (or substitutional solute) diffusional phenomena and related segregation or compound formation (precipitation). This is particularly true for variations in diffusivity provided by interfaces such as grain boundaries or phase boundaries as well as dislocations. These variations in

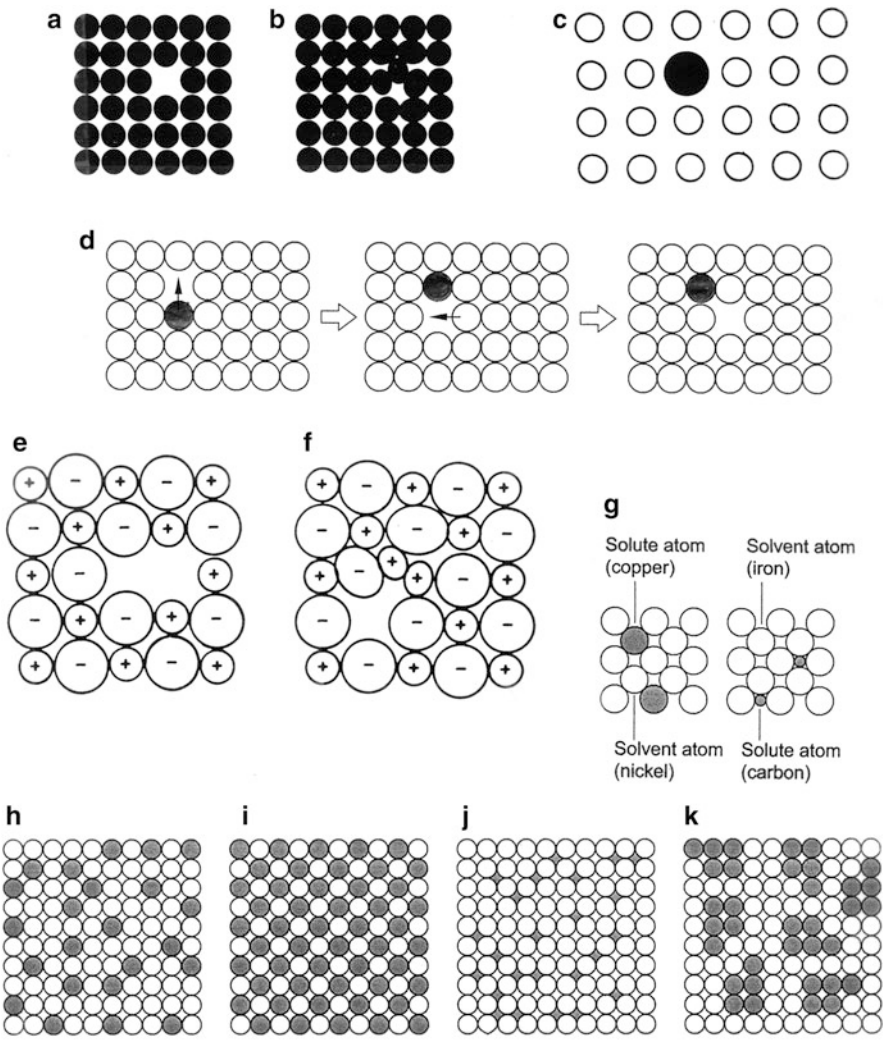


Fig. 1 Point defects in crystal lattices. (**a**) vacancy, (**b**) interstitial, (**c**) substitutional impurity, (**d**) vacancy-substitutional impurity diffusion, (**e**) Schottky defect, (**f**) Frenkel defect, (**g**) solute-solvent lattice conversion, (**h**) random, (**i**) ordered, (**j**) random interstitial, (**k**) solute clusters in solid solution

diffusivity, D, are typically illustrated in Fig. 2, and diffusivity (or diffusion coefficient) is represented generally by

$$D = D_o e^{(-Q/RT)} \tag{3}$$

where Q is the activation energy for the particular transport process shown in Fig. 2b and R is the molar gas law constant, while the transport or flux (J) of diffusing species is expressed by *Fick's first law*:

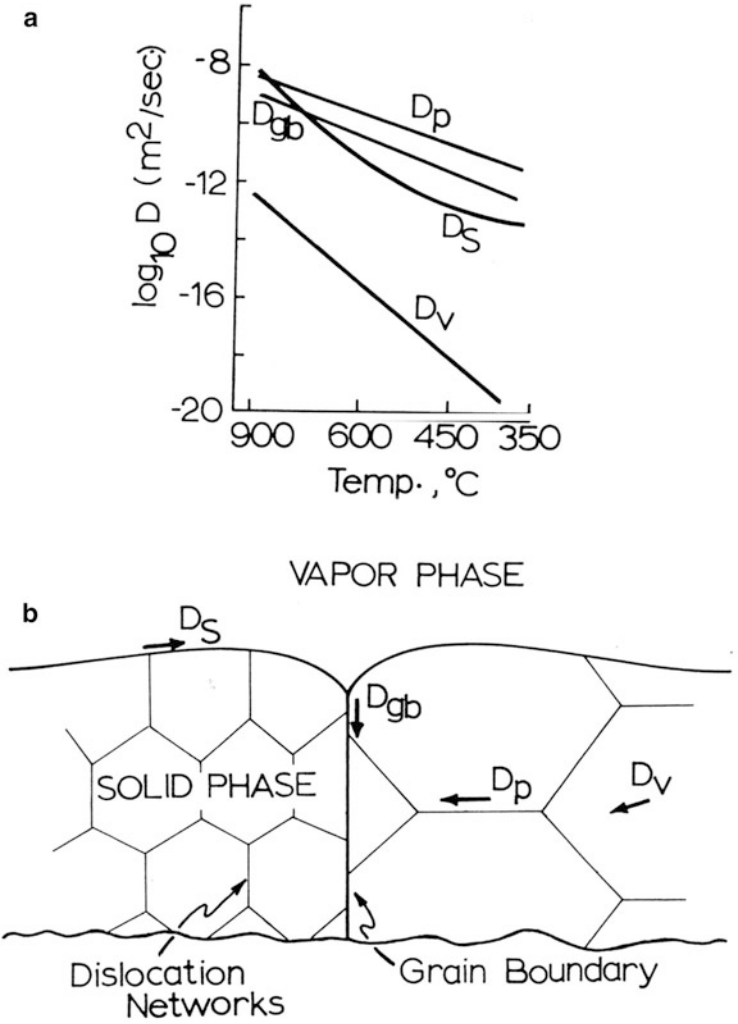


Fig. 2 Variation of diffusion coefficients with temperature in polycrystalline Ag (**a**). (**b**) Shows the corresponding diffusion paths or diffusion modes: D_s, surface diffusion (diffusion coefficient); D_{gb}, grain boundary diffusion; D_p, dislocation (pipe) diffusion; D_v, volume diffusion (Adapted from Murr (1975))

$$J = -D(dC/dx) \tag{4}$$

and C is the concentration of the diffusing species in the x-direction: dC/dx therefore represents the concentration gradient. The variation of concentration with time is expressed by *Fick's second law* of diffusion:

$$\frac{dC}{dt} = D\frac{\partial^2 C}{\partial x^2} \tag{5}$$

There are many examples of interstitial (solute) diffusional phenomena in materials systems. The iron-carbon system represented in Fig. 1g is a case in point. There are two important examples of interstitial carbon diffusion in iron or iron alloys, especially involving dislocation diffusion. These will be described in the following.

Low-Carbon Steel in the Caribbean Steel Drum: The Role of Interstitial Carbon

Following the Second World War, the beaches of the island of Trinidad were strewn with 55-gal oil barrels. This provided the native Trinidadians the opportunity to configure empty barrels into tonal percussion instruments, originally with 14 musical notes raised on a concave, domed platform on one end of the drum. Later drums evolved into a variety of orchestral voices, with soprano or lead drums containing from 29 to 31 notes configured into three octave ranges. These instruments, the only truly musical instrument developed in the twentieth century, are not really drums since the notes are ideally elliptical shells and emulate bells, having chromatic overtunes. The instrument is popularly referred to as a pan, and each note is tuned to a fundamental frequency (f) in the note center, with the long or major axis of an elliptical note tuned to the first octave (2f) at its ends and at least one or more harmonic at the ends of the short or minor axis (3f, 4f (the second octave), etc.). This produces chromatic tones for each note and sympathetic resonance of the note harmonics. Figure 3 illustrates the Caribbean drum (or pan) where small amplitude modal frequencies can be expressed by

$$f_n = \frac{\omega_n}{2\pi}\left(\frac{h}{a^2}\right)\sqrt{\frac{E}{\rho(1-\nu)}} \tag{6}$$

The dimensionless eigenfrequency term, $\omega_n = n\omega$, $=> nf_1$, where f_1 is the fundamental frequency or note tonic. The note thickness is denoted h, while the note size (approximated as a circular note), or radius, is denoted by a in Fig. 3a. E, ρ, and υ in Eq. 6 are the Young's modulus, density, and Poisson ratio, respectively. The instruments illustrated in Fig. 3a, d have the barrel cut away leaving what is called a "skirt." It was found that sound amplification could be optimized by cutting shorter skirts for higher-frequency-range drums in contrast to lower-frequency-range drums, with the soprano (or lead) drum having the shortest skirt in contrast to a base drum which retains the full barrel length (Fig. 3e). The base drum has only three large notes on a single drum platform, but six such platforms compose the base ensemble (Fig. 3e).

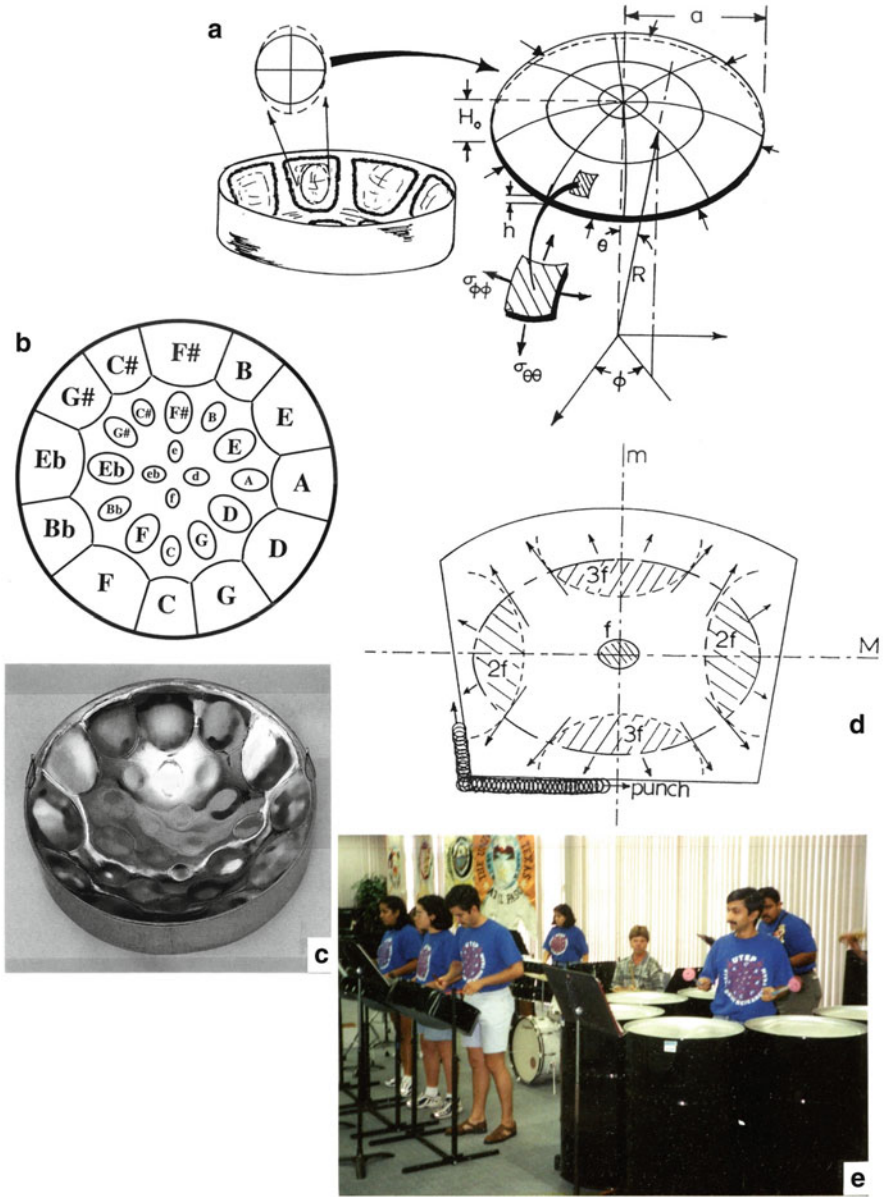


Fig. 3 Caribbean steel drum elements and configurations. (**a**) Steel drum and note schematic. The note rise height is denoted H_o. (**b**) Steel drum note (tonal) pattern. (**c**) Chrome-plated (for aesthetic appearance) soprano or lead drum patterned as in (**c**). (**d**) Frequency zones for tuned, elliptical note. The note tonic or fundamental is f (or f_1). (**e**) Playing steel drums

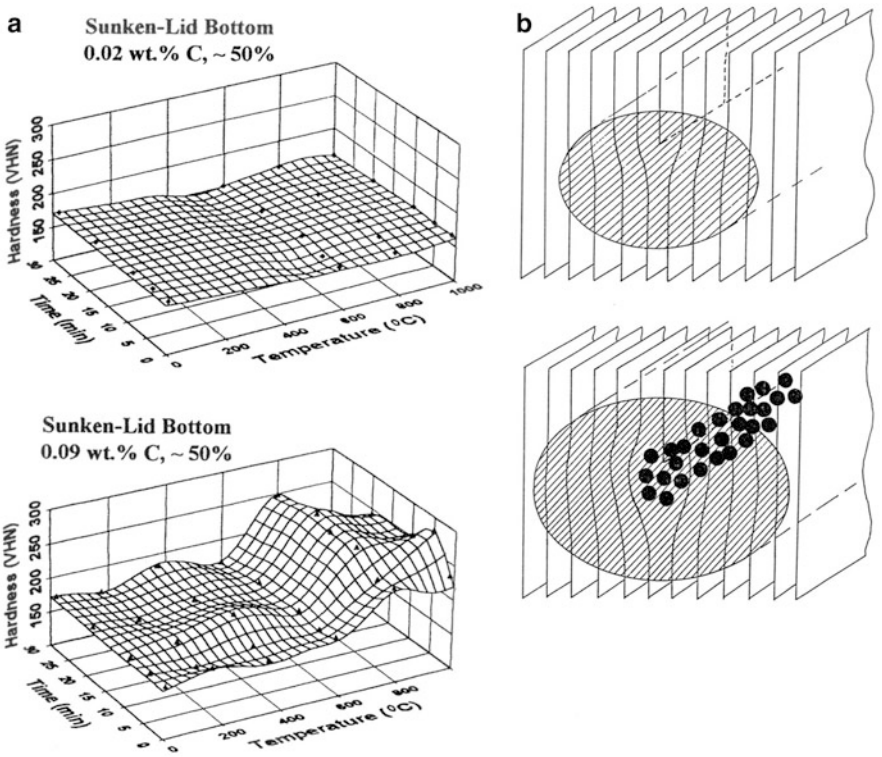


Fig. 4 (**a**) Examples of tempering diagrams for the bottom of two steel drum domes with different carbon contents. The deformation is 50 % thickness reduction. (**b**) Edge dislocation before and after interstitial carbon diffusion along the strained core region, creating a Cottrell "atmosphere" which pins the dislocation, will also increase the sensitization (or decrease the time to sensitize, and form $Cr_{23}C_6$ precipitates). This occurs because the grain boundary area increases with decreasing grain size, and dislocation pipes or dislocation density increase with increasing strain and strain state (ϵn). Fig. 6 illustrates the variation of sensitization for different grain sizes of 304 stainless steel subjected to varying tensile strains and aged at 625 °C, which provides some quantification for Fig. 5b

Early on, it was noticed that only after burning the paint and oil from configured drums by positioning them over a fire (often a burning tire which matched the drum diameter) could the instrument be adequately or optimally tuned. It was also noticed that the time of heating made a difference, but in some instances, note-configured instruments could not be tuned.

It was not until the mid-1990s that the Caribbean steel drum was studied as a materials system (Murr et al 1999; Ferreyra et al. 1999), and in 2001, Ferreyra et al. (2001) demonstrated that the heating of the configured instrument involved tempering or strain aging. In addition, aging only occurred for carbon content above 0.02 % (by weight) and overaged rapidly for carbon content above ~0.1 %. Figure 4a shows

representative tempering (aging) diagrams for instrument platforms formed from two different carbon-content barrels, while Fig. 4b illustrates the diffusion of interstitial carbon to dislocation cores during aging, creating what is historically referred to as a Cottrell (solute) "atmosphere," which acts to pin the dislocation. Under tensile deformation, this would cause discontinuous or serrated yield phenomena as dislocations are pinned and unpinned with the applied strain (Cottrell 1953).

It might be noted that interstitial carbon, in contrast to silver in Fig. 2a, diffuses roughly four orders of magnitude faster, especially along dislocation cores (or "pipes"). As shown in Fig. 4a, the bottom of the domed platform, where the reduction in thickness is 50 %, and for a carbon content of 0.09 %, strain aging is optimum at ~10 min at 800 °C, where the base hardness is elevated from ~VHH 175 (1.75 GPa) to ~VHH 250 (2.5 GPa). This hardness and corresponding carbon-diffused dislocation structures create optimal chromatic tonal qualities in the tuned instrument. In contrast, the drum head with 0.02 % carbon exhibits no aging (Fig. 4a).

It should be noted that doming or sinking of the 55-gal barrel lid is done by hand, with a 6-lb sledge hammer; forming a nearly perfect hemispherical or near hemispherical platform. Notes are positioned and punched up from the back side, and as illustrated schematically in Fig. 3c, d. Notes near the rim, where the lid thickness is deformed by only about 10 % (10 % reduction in thickness), are of lower frequency and larger (with larger radii, a, in Fig. 2a) in contrast to the highest frequency and smaller notes at the bottom of the domed platform where the reduction is 50 % and the note thickness h in Fig. 3a is smaller (Eq. 6).

Precipitation and Sensitization in Austenitic Stainless Steels: The Role of Interstitial Carbon

In the decades following the installation of stainless steel piping in boiling water reactors and similar applications, severe intergranular stress corrosion cracking (IGSCC) was observed (Klepter 1975; MacDonald 1992), especially in 300 series stainless steels containing sufficient chromium to passivate corrosion especially at grain boundaries. Subsequent investigations of the underlying fundamentals revealed that piping with small grain sizes or heavily deformed, mill-produced piping, or both, were particularly prone to corrosion sensitization and eventual grain boundary corrosion. Sensitization, especially grain boundary sensitization, involves the diffusion of carbon atoms to the grain boundaries where nucleation is energetically favorable for the formation of $M_{23}C_6$ precipitates (M being primarily Cr). As these precipitates are heterogeneously nucleated and grown, the Cr in the grain boundary is depleted below the passivation threshold of ~13 wt%, and corrosion commences. Since, as illustrated ideally in Fig. 2, both grain boundary and dislocation pipe diffusion (Fig. 4b) are more rapid than the matrix diffusion, especially for interstitial carbon, smaller grain size and higher dislocation density will both accelerate the rate of interstitial carbon transport and accelerate both sensitization (Cr depletion) and IGSCC. As the carbon content increases in the grain boundaries, the Cr content decreases, and the corresponding thermo-kinetics are illustrated in a typical

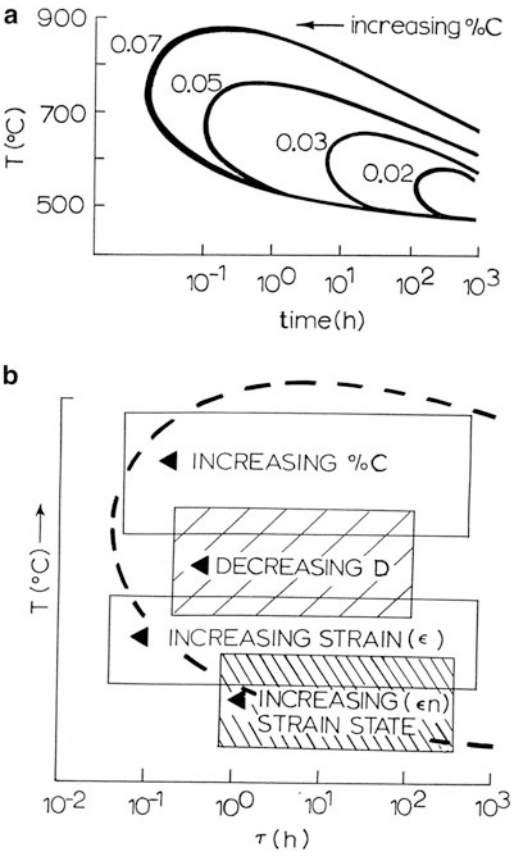


Fig. 5 (**a**) Time-temperature-sensitization curves for 304 stainless steel with varying carbon contents shown. $Cr_{23}C_6$ precipitation occurs in the areas to the right (as shaded for the lowest carbon content). (**b**) Combined effects of carbon content (%C), grain size (D), strain (ϵ), and strain state (ϵ(n)) on time to sensitize (After Trillo et al. (1995))

time-temperature-sensitization (TTS) diagram shown in Fig. 5a for 304 stainless steel (~18 % Cr, 8.5 % Ni, balance Fe in weight percent) having a constant grain size and corresponding residual strain, and different interstitial carbon content (by weight percent). It can be observed in Fig. 5a that the aging time for carbide ($Cr_{23}C_6$) precipitation is increased significantly as the carbon content is decreased but, as shown qualitatively in Fig. 5b, decreasing grain size and increasing strain and strain state (uniaxial, biaxial, or triaxial; $n = 1, 2, 3$, respectively) (Fig. 6).

Since the heterogeneous nucleation which occurs in the grain boundaries is expedited by the available interfacial free energy, the rate of sensitization will also depend largely on the boundary energy. This is very convincingly illustrated in Fig. 7a which clearly illustrates $Cr_{23}C_6$ precipitates forming on regular grain boundaries (gb) and non-coherent twin boundary steps (TB) with average interfacial free energies of ~1,026 mJ/m^2 and 265 mJ/m^2, respectively, (Fig. 7c) after aging for 50 h at 620 °C. Note that there are no indications of precipitates on the coherent {111} twin boundary portions (tb) (Fig. 7a) where the corresponding interfacial free energy is only about 16 mJ/m^2. Figure 7b shows an inclined grain boundary portion with some indication of precipitation along grain boundary ledge

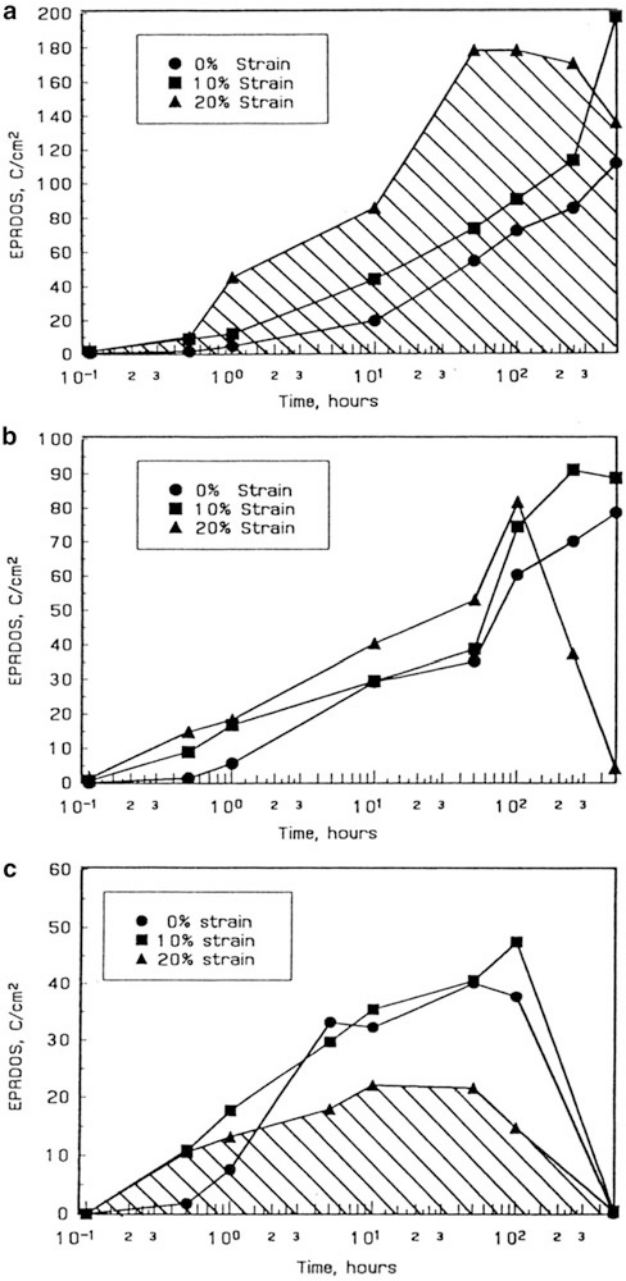


Fig. 6 Comparison of EPR-DOS (Electrochemical Potentiostatic Reactivation-Degree of Sensitization) curves for type 304 stainless steel with 0.051 % C deformed in tension at strains shown (10^{-3} s^{-1} strain rate) and aged at 625 °C: (**a**) 150 μm grain size, (**b**) 40 μm grain size, (**c**) 15 μm grain size (After Trillo et al. (1995))

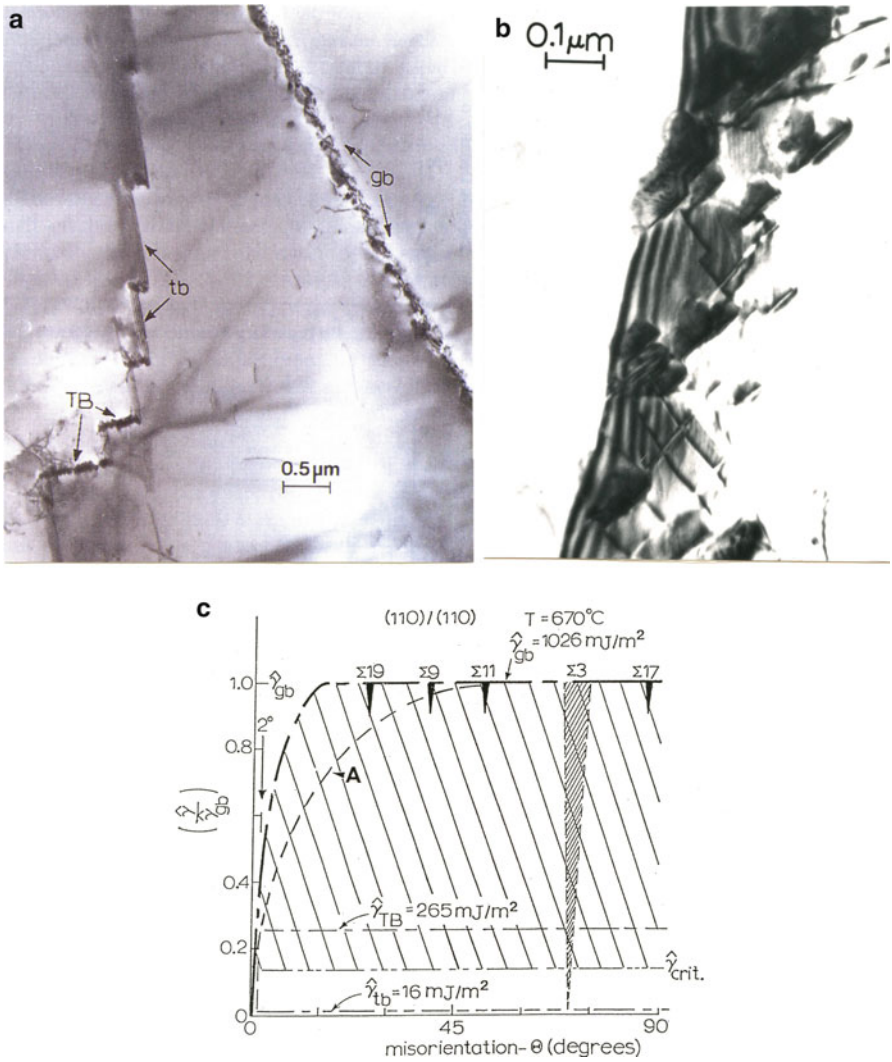


Fig. 7 (**a**) TEM bright-field image showing precipitation of $Cr_{23}C_6$ precipitates at interfaces in unstrained 304 stainless steel having a grain size of 50 μm and 0.07 % C, aged at 670 °C for 50 h. (*gb*) represents a high-angle grain boundary, (*tb*) coherent {111} twin boundary, and (*TB*) non-coherent twin boundary steps. (**b**) Inclined grain boundary portion showing precipitates. (**c**) Schematic representation of energy versus boundary misorientation. Boundary energies correspond to 670 °C. Grain orientations are (110). The *arrow* at 2° indicates the lower misorientation angle where precipitation of carbides was observed. (**a**) denotes the actual energy versus misorientation measured. Cusps at $\sum$ 19, 9, 11, and 3 indicate coincidence boundaries. $\sum$ 3 represents the coherent {111} twin boundaries (Adapted From Trillo and Murr (1995))

structures, where the energy would also be optimized in the grain boundary plane as discussed in more detail later in this chapter.

In addition to the Cr depletion which eliminates grain boundary passivation and allows for corrosion in appropriate environments, the carbide precipitates which grow in the boundaries as illustrated in Fig. 7a, b promote cracking and grain boundary embrittlement. Obviously, engineering strategies to reduce or eliminate these phenomena can include reduced carbon content, increased grain size, reduced internal strain (or dislocation density) (Fig. 5), or, more importantly, engineering the boundary type, increasing the interfacial fraction of coherent twin boundaries (Fig. 7a, c), thereby retarding or eliminating sensitization and corrosion as implicit in Fig. 7a.

References

Cottrell AH (1953) The Portevin-Le Chatelier effect. Philos Mag 44:829–832

Ferreyra E, Maldonado JG, Murr LE, Papper S, Kennedy C, Posada M, DeAlba J, Chitre R, Russell DP (1999) Materials science and metallurgy of the Caribbean steel drum. Part II. Heat treatment, microstructures, hardness profiles, and tuning effects. J Mater Sci 34:981–996

Ferreyra E, Murr LE, Russell DP, Bingert JF (2001) Elastic interactions and the metallurgical and acoustic effects of carbon in the Caribbean steel drum. Mater Charact 47:325–363

Klepter HH (1975) Investigation of the cause of cracking in austenitic stainless steel piping. General electric report NEDO – 21000

Macdonald D (1992) Corrosion in boiling water nuclear reactors. Earth Miner Sci 61:27–32

Murr LE (1975) Interfacial phenomena in metals and alloys. Addison-Wesley Publishing co, Reading, Ma

Murr LE, Ferreyra E, Maldonado JG, Trillo EA, Papper S, Kennedy C, DeAlba J, Posada M, Russell DP, White JC (1999) Materials science and metallurgy of the Caribbean steel drum: part I. Fabrication, deformation phenomena, and acoustic fundamentals. J Mater Sci 34:967–979

Trillo EA, Murr LE (1995) Effects of carbon content, deformation, and interfacial energetic on carbide precipitation and corrosion sensitization in 304 stainless steel. Act Mater 47 (1):235–245

Trillo EA, Beltran R, Maldonado JG, Romero RJ, Murr LE, Fisher WW, Aduani AH (1995) Combined effects of deformation (strain and strain state), grain size, and carbon content on carbide precipitation and corrosion sensitization in 304 stainless steel. Mater Charact 35:99–112

Line Defects: Dislocations in Crystalline Materials

Contents

Abstract

Dislocation geometries, energetics, generation, interaction, dissociation, and multiplication are described as these relate to deformation and related properties in crystalline materials. Rotation-induced imperfections described as disclinations are also discussed.

Introduction

Line defects, especially edge and screw dislocations, are primarily responsible for characterizing slip which accommodates shear-resolved deformation. In contrast to point defects, which might be considered zero-dimensional or 0-D, line defects might be considered 1D. Dislocations arise at interfacial (and surface) irregularities. These irregularities constitute sources or source mechanisms which are activated by a resolved stress (or strain) and are demarcations for slip in energetically appropriate crystal planes: {111} in fcc materials, {110} in bcc materials, etc., representing close packing of the atomic structure. They are characterized by a corresponding line energy, and as they are created and multiply by interactions with one another, the energy implicit in deformation is manifested in their density: number of defects/unit volume or the number of dislocation line intersections with some reference surface area (number/unit area).

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_15

In face-centered cubic (fcc) materials in particular, total or primary dislocations are often observed to split into two secondary or partial dislocations in order to reduce their energy. They are connected on their respective {111} slip plane by an interfacial irregularity, a stacking fault, which also possesses some energy/unit area. However, this combined entity of dissociation is overall less than that possessed by a total or primary dislocation.

This chapter will explore the fundamental properties of dislocations or line defects in crystalline materials, including their crystallographic features which are manifest in a defining displacement vector called the Burgers vector. As it turns out, simple derivations will show that the dislocation line energy is proportional to the scalar value of the Burgers vector squared. And because the level of deformation imparted in a material is related to the energy expended in the process, the dislocation density is a reflection of the relative level of deformation. These issues will be illustrated using transmission electron microscope (TEM) observations of dislocations, dislocation structures, and dislocation densities corresponding to deformation processes.

Volterra Dislocations

While dislocations were originally postulated by G.I. Taylor, E. Orowan, and M. Polanyi around 1934 to account for deformation-related displacement or slip in crystal lattices, V. Volterra, nearly 30 years prior, considered the more general elastic properties of a cut cylinder to be "deformed" in various ways to produce six types of dislocations. Figure 1 illustrates the Volterra dislocations, also referred to as Volterra distortions. Figure 1b, c represents displacements characterizing edge dislocations, while Fig. 1d corresponds to the screw dislocation. Figure 1e–g represents displacements which are proportional to the outer cylinder radius and is nonvanishing as the radius tends to infinity. These dislocations are rotation induced and characterized by a relative displacement produced by a mutual rotation by an angle ω around a fixed axis. These dislocations are generally not crystalline imperfections and have come to be called *disclinations*, where, in contrast to the classical edge and screw dislocations in Fig. 1b, d, whose strength is determined by the Burgers vector, **b**, or the corresponding transitional displacement, the disclination strength is related to the axial vector, ω (referred to as the *Frank vector* (Frank 1958)), which defines the mutual rotation or relative displacement in Fig. 1e–g in the form (Romonov and Kolesnikova 2009)

$$\mathbf{u} = \boldsymbol{\omega} \times (\mathbf{r}_{\omega} - \mathbf{r}), \tag{1}$$

where $\boldsymbol{\omega}$ is the axial Frank vector, $\mathbf{r}$ is the radius vector, and $\mathbf{r}_{\omega}$ is the position of the axis for $\boldsymbol{\omega}$. Equation 1 represents a displacement corresponding to a rigid body rotation. While both dislocation Burgers vector at dislocation Frank vector magnitudes must obey crystal symmetry considerations when applied to crystalline materials, disclinations have long-range distortion fields and are a source of large

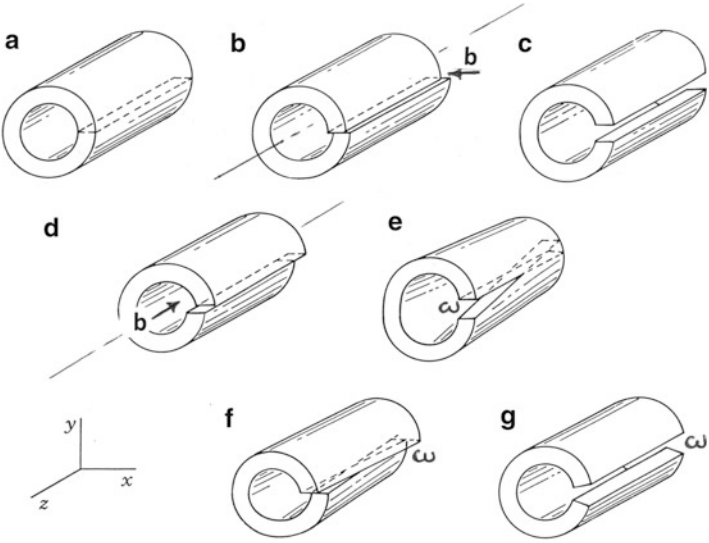


Fig. 1 Volterra dislocations or topological defects (Volterra 1907). (**a**) Basic cut cylinder. (**b**) and (**c**) are edge dislocations resulting by linear displacements. (**d**) is the screw dislocation characterized by a linear displacement (**b**). (**e**–**g**) are disclinations characterized by rotational displacements (ω) (Adapted from Hirth and Lothe (1968))

internal elastic stresses. Consequently they possess a very large energy in comparison with edge or screw dislocations and are thermodynamically only tractable on considering higher energy imperfections such as dislocation configurations at grain boundaries or special deformation-induced defects such as microbands and nanoparticles or nanoscale microstructures. Disclinations also provide topological constructions or line defects in liquid crystal systems. F. R. H. Nabarro (1967) not only described some of these disclination features in crystals but also proposed disclination motion as a major contribution to muscle contraction which involves large molecular structure displacements (Nabarro 1969).

Figure 2a, b illustrates perfect disclinations in crystalline structures as envisioned by Nabarro (1967), while Fig. 2c shows the relationship between wedge disclinations and terminated dislocation walls characteristic of a grain boundary (Kim 2010). In this regard, the wedge disclination shown in Fig. 2b is observed to be a pentagon inserted into the graphene or hexagonal carbon lattice or lattice fragments shown in ► Figs. 5a of chapter "Examples of Materials Science and Engineering in Antiquity" and ► 11 of chapter "Chemical Forces: Nanoparticles", where the pentagonal defect creates curvature in the planar graphene fragment. Figure 2d, e shows so-called positively and negatively charged disclinations (Nabarro wedge disclination in Fig. 2b), and their combination to create special edge dislocations in the graphene lattice to form a symmetrical grain boundary is represented at the right in Fig. 2c. Note that the equivalent edge dislocation at the left in Fig. 2e represents a special wedge disclination which forms a perpetual grain boundary away shown to the right in Fig. 2e.

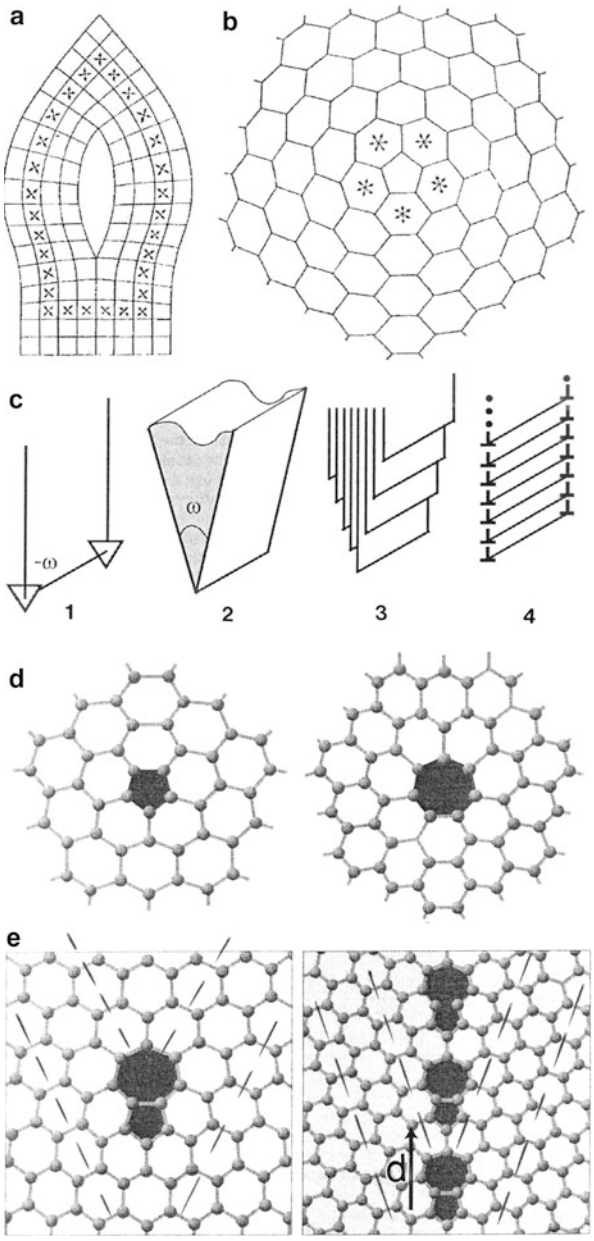


Fig. 2 Wedge disclination examples. (**a**) Positive wedge disclination of strength π/2. (**b**) Positive wedge disclination of strength π/3; ω = 2π/n, where n = 2, 3, 4, 6 (Nabarro 1967). (**c**) shows relationship between wedge disclination and terminated edge dislocation walls characterizing a grain boundary: (1) negative wedge disclination with strength – ω, (2) wedge of angle ω, (3) half planes modeling the wedge in (2), (4) equidistant edge dislocation wall (After Hirth (1963)). (**d**) Positively (*left*) and negatively (*right*) charged disclinations as topological disorder in graphene. (**e**) Equivalent edge dislocation (*left*) and continuous grain boundary (*right*) formed by periodic array of these equivalent edge dislocations with a periodicity defined by d (Adapted from Kim (2010))

It is interesting to note that while wedge disclination structures shown in Fig. 2d, e can be considered topological disorders in graphene, graphene itself in carbon onions and multiwalled carbon nanotube structures shown in ▶ Figs. 13b and ▶ 19 of chapter "Chemical Forces: Nanoparticles" can be considered an edge dislocation topological defect especially implicit in the large carbon onion structure shown in Fig. 3a.

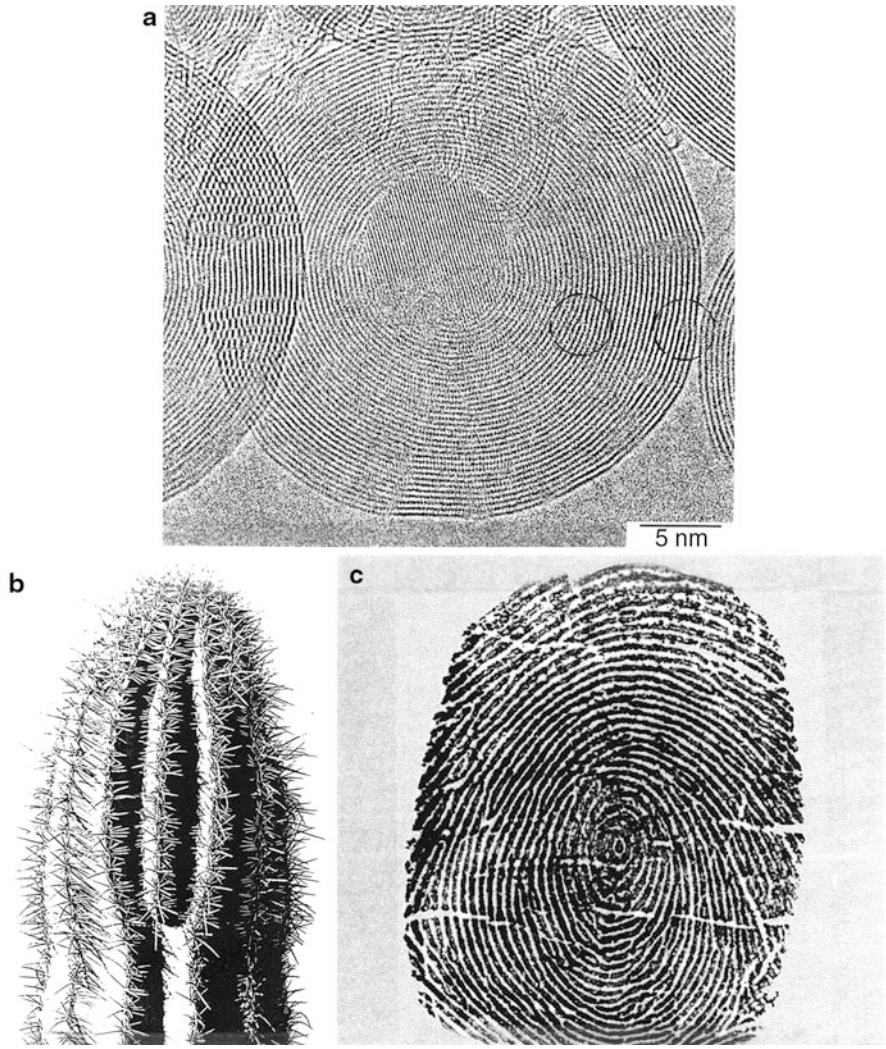


Fig. 3 Edge dislocation topological defects. (**a**) Edge dislocations in carbon onion. The *center* crystalline region is diamond cubic structure (Adapted from Banhart and Ajayan (1996)). (**b**) Edge dislocation in Saguaro cactus. (**c**) Edge dislocations in fingerprint. These are also characteristic of partial disclination dipoles

Topological continuity is also provided by edge dislocations in relaxing curvature by pleats or so-called topological dipoles as shown in Fig. 3b, c for a desert cactus and a fingerprint, respectively.

It might also be expected that since, as demonstrated in Fig. 2, pentagonal defects are characteristically wedge disclinations, they would play a role in quasicrystals. These phenomena have in fact been discussed by a number of authors (Bohsung and Trebin 1987; Leung et al. 1989).

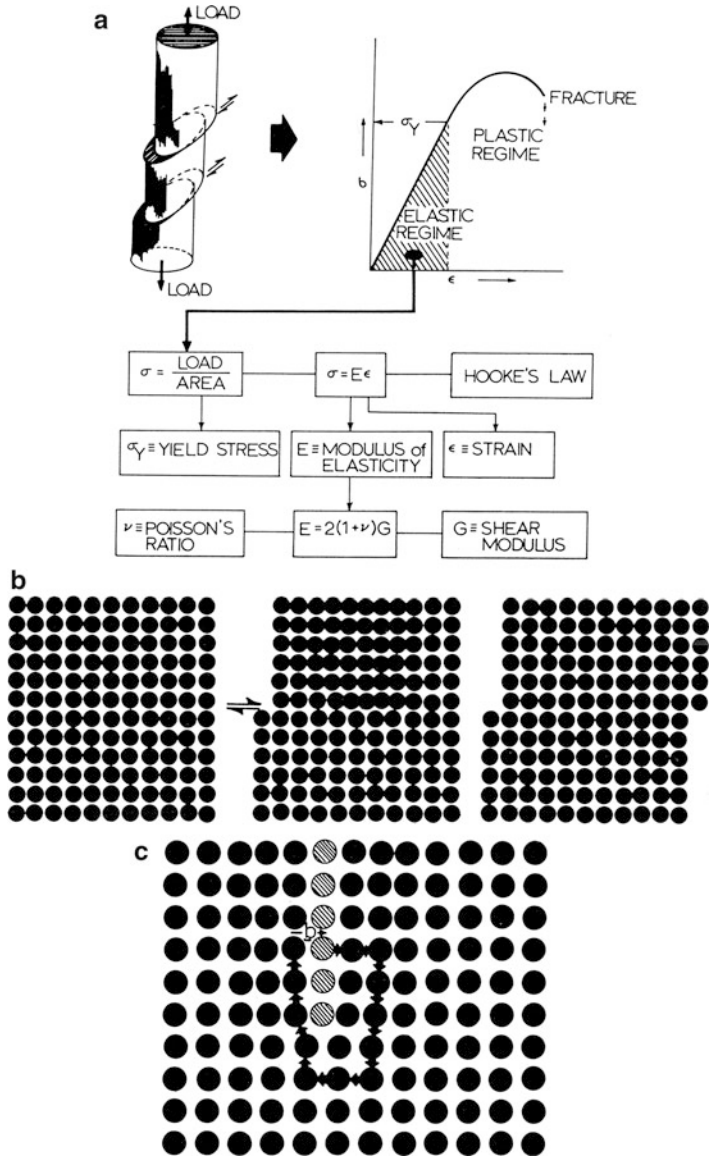


Fig. 4 Slip deformation and edge dislocations. (**a**) Tensile deformation involving planar slip (shear deformation) and the corresponding stress (σ) – strain (ϵ) diagram with defining parameters. (**b**) Shear-induced slip by unit displacement forming an edge dislocation which emerges to form a step and unit displacement of the crystal portion relative to another. (**c**) Edge dislocation with Burgers circuit and Burgers vector, b

Figure 4b illustrates an edge dislocation represented by an extra half plane whose edge defines the dislocation "line" and the slip plane on which this extra half plane can be translated under an applied (or resolved) shear stress. The shaded

region in Fig. 4b (top) represents the strain field which results by crystal plane bending around the dislocation line or plane edge. Figure 4a represents tensile deformation-induced slip (translation) on oriented shear planes, while Fig. 4b illustrates the unit displacement creating an edge dislocation which propagates through the crystal to create a unit displacement (shear) of one crystal portion relative to the other. Figure 4c shows the unit displacement by a closure failure of a Burgers circuit around the edge dislocation represented by the shaded atoms composing the extra half plane. This closure failure represents the displacement creating the dislocation in Fig. 4b and is called the Burgers vector, **b**, since this displacement vector is perpendicular to the dislocation line vector (ξ) which is perpendicular to the page in Fig. 4c, and it defines the edge dislocation ($\mathbf{b} \perp \xi$).

Figure 5a illustrates a total dislocation in a cubic crystal section which shows the edge dislocation in the front face and becoming a screw dislocation on the side face. The edge and screw dislocation line vectors, ξ, are perpendicular to one another, while for the screw dislocation line vector, the Burgers vector is parallel to it ($\mathbf{b} \| \boldsymbol{\zeta}$). The dislocation line is represented at the edge of the shaded region in Fig. 5a, and its character (or the character angle between the dislocation line vector, $\boldsymbol{\zeta}$, and the Burgers vector) becomes mixed between the edge character (90°) in the front face and the screw character (0°) in the side face. The shaded region can be considered a cut through the corner or edge of the section and then displaced (**b**) perpendicular to the edge dislocation line as shown in Fig. 5a.

Dislocation Line Energy (Self-Energy)

In Fig. 3b, the dislocation line vector, $\boldsymbol{\zeta}$, in Fig. 3a is considered to lie along the z-axis in a cylindrical section having a diameter R, extracted from the crystal. In the one case, the z-axis is oriented along the screw dislocation line vector, $\boldsymbol{\zeta}$, while in the other it is oriented along the corresponding edge dislocation line vector. This cylindrical section is cut into the cylinder center. For the screw dislocation, the top cut is displaced an amount **b** parallel to the dislocation line vector (or z-axis), while for the edge dislocation, the top of the cut is displaced an amount **b** perpendicular to the dislocation line vector (also the z-axis). Utilizing polar coordinates as shown for the screw dislocation, the stress associated with the displacement, **b**, parallel to the dislocation line vector, $\boldsymbol{\zeta}$, is

$$\sigma_{\theta z} = \frac{Gb}{2\pi r}, \tag{2}$$

where G is the shear modulus represented in Fig. 4a. The elastic strain energy stored in a unit length of the screw dislocation associated with the stress field of an infinite dislocation in the bounded cylinder in Fig. 5b is expressed by

$$\xi\,(\text{screw}) = \int_{r_0}^{R} \left(\frac{\sigma_{\theta z}^2}{2G}\right) 2\pi r dr, \tag{3}$$

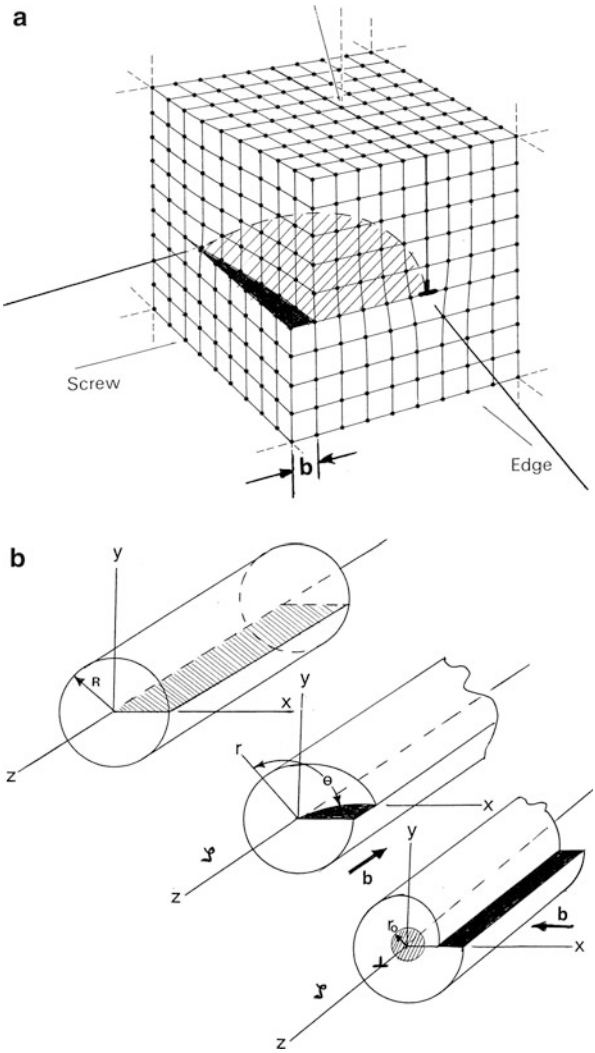


Fig. 5 Total (edge and screw) dislocation convention in a crystal section. *Shaded region* shows cut, while solid portion is the displacement of *top* section of cut relative to the *bottom*. *b* shows the displacement (Burgers) vector. (**b**) Solid cylinder with cut into cylinder axis showing displacement parallel to the axis (screw dislocation) and perpendicular to the axis (edge dislocation)

$$\xi\,(\text{screw}) = \frac{\text{Gb}^2}{4\pi}\ell\text{n}\left(\frac{\text{R}}{r_0}\right), \tag{4}$$

The rationale for elastic stored energy is implicit from the stress–strain diagram in Fig. 4a where it can be observed that the elastic strain energy (U_E) represented by the shaded portion (elastic regime) is given by

$$U_E = \sigma_y \epsilon_y/2 = \sigma_y^2/2E = \sigma_y^2/4(1+\upsilon)G, \tag{5}$$

where σ_y is the ideal yield stress.

Correspondingly, for the edge dislocation in Fig. 5b, the only nonzero stress is given by

$$\sigma_{xy} = -\frac{Gb}{2\pi(1-\upsilon)x}, \tag{6}$$

where υ is Poisson's ratio represented in Fig. 4a. Similarly, the elastic strain energy stored in a unit length of the edge dislocation can be expressed by

$$\xi(\text{edge}) = \int_{x=r_0}^{x=R} \left(\frac{\sigma_{xy}^2}{2G}\right) 2\pi x dx \tag{7}$$

$$\xi(\text{edge}) = \frac{Gb^2}{4\pi(1-\upsilon)} \ln\left(\frac{R}{r_0}\right) \tag{8}$$

The value of r_0 in Eqs. 4 and 8 is considered to represent the strain field radius associated ideally with an edge dislocation as represented in the shaded area of Figs. 4b and 5b. This is usually approximated by $r_0 \cong 5b$. While for many engineering metals and alloys, $\upsilon \cong 1/3$, and the associated edge and screw dislocation line energies would differ by ~3/2. The important feature of dislocation line or self-energy in that it is proportional to the Burgers vector squared: $\xi \alpha b^2$.

Since the dislocation line vector, ξ, is in the slip plane, which is crystallographic, the Burgers vector is also in the slip plane, and also crystallographic, a crystal direction. For fcc crystals, the slip plane is always {111}, and dislocation Burgers vectors are of the form $a < 110 > /2$, where a is the lattice parameter. For bcc crystals, the slip plane is normally {110} and the Burgers vector is of the form $a <111>/2$ for the smallest, perfect (or total) dislocation. Since any slip plane or slip system in bcc which contains this crystallographic direction can be a potential slip plane, {112} and {123} can act as slip planes, and slip has in fact been observed in these planes.

In ionic crystals with the NaCl-type structure (▶ Fig. 2 of chapter "Electrovalent Crystal Structures and Chemistry"), dislocations lie in {110} planes. Because dislocation lines meeting the surface of an NaCl-type single crystal will have an associated self-energy (Eqs. 4 and 8), appropriate and selective etching of crystals containing dislocations exhibits selective etching along dislocation lines meeting the surface, creating crystallographic etch pits as illustrated for dislocations in (001) LiF in Fig. 6a and shown schematically in Fig. 6b. Early indications of dislocations were observed in NaCl-type crystals (Gilman et al. 1958), and Gilman (1969) studied dislocation dynamics by observing the motion of dislocations or their displacement from etch pits as shown by the flat-bottomed pits in the upper right in Fig. 6a. This occurs, as shown in Fig. 6b, by the dislocation moving after a pit forms, and the absence of the dislocation line allows the pit to bottom out and its base to become flat. Dislocations were also observed by transmission electron microscopy about the same time, which allowed the dislocation line itself to be observed inside the crystal and to be characterized crystallography, including its Burgers vector.

Figure 7 shows edge dislocations observed in high-resolution transmission electron microscope (TEM) images of crystal lattice (atomic) planes. The $Cr_{23}C_6$

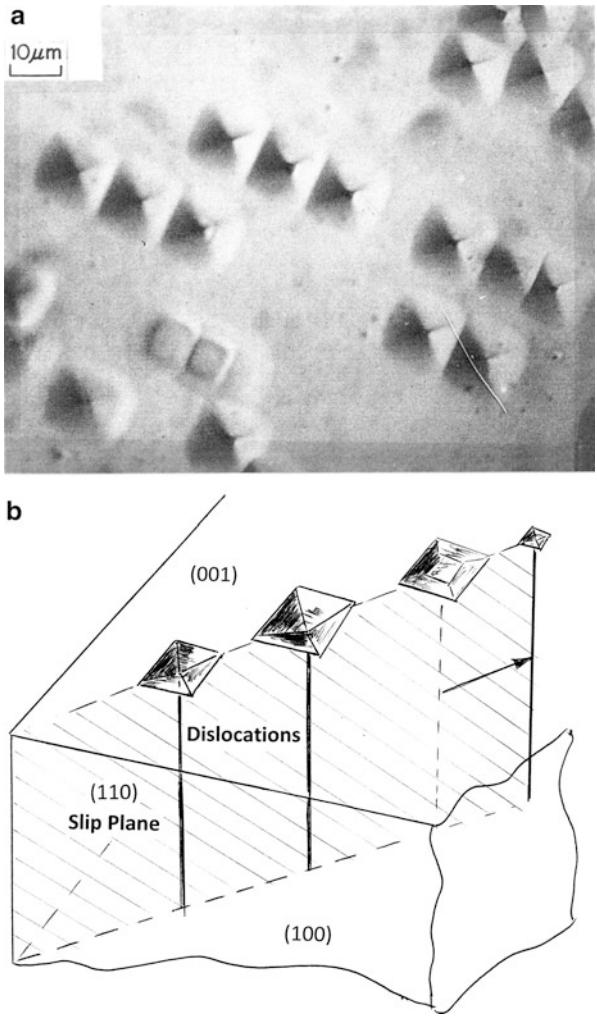


Fig. 6 Dislocation etch pits in LiF (**a**). The (001) face was etched in concentrated HF containing powered Fe as a reaction catalyst. (**b**) shows the dislocations on the (110) plane and the etching of pyramidal (inverted) pits along the dislocation line. The pit bottom flattens when the dislocation moves away

precipitates shown in Fig. 7a–c are characteristic of those shown in the grain boundary regions in Fig. 7a, b. The dislocation in a palladium thin film shown within the dotted circle in Fig. 7d can be readily observed by viewing into the page at a shallow angle along the crystal lattice planes.

Dislocation Production, Dissociation, Interaction, and Multiplication

In a perfect crystal, or a single crystal, ideally containing no dislocations, only the free surface can produce (or nucleate) dislocations and only at elastic incompatibilities which can exist at surface steps or ledges as shown in Fig. 8a, b. Even in the unstrained condition an atomic step can be observed computationally to exhibit a

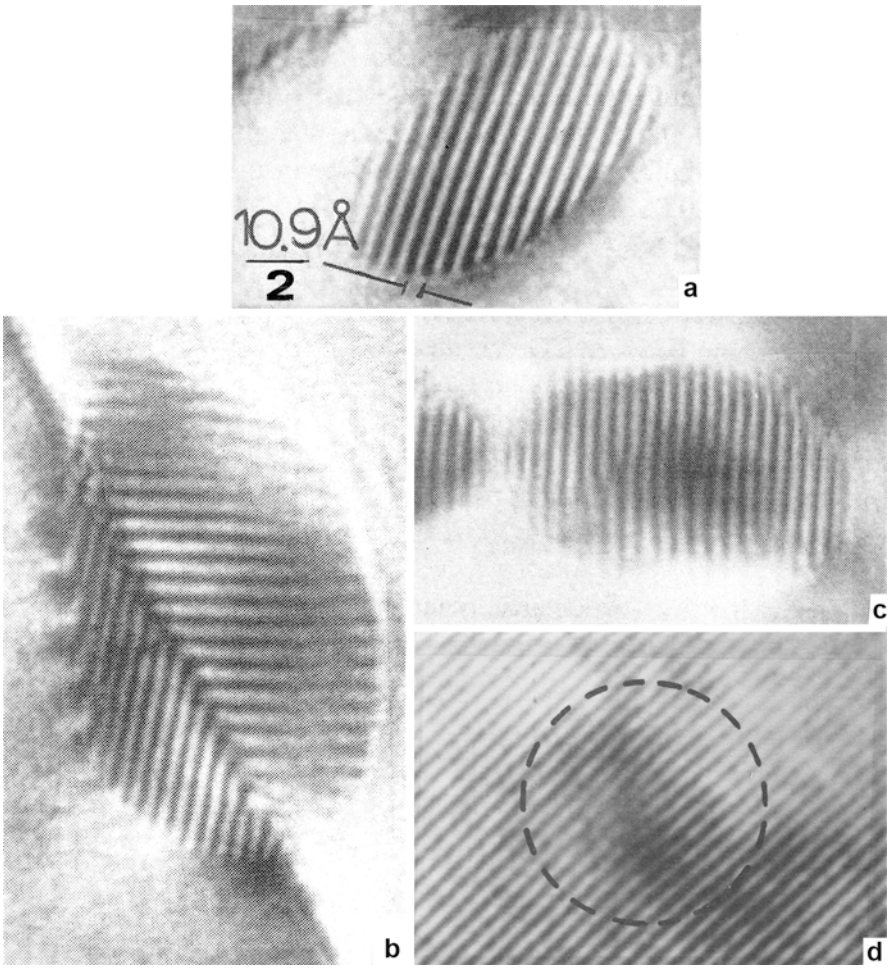


Fig. 7 Edge dislocation crystalline structures. (**a**–**c**) show $Cr_{23}C_6$ single-crystal precipitates observed at high resolution in the TEM. (**a**) Lattice fringes show (200) planes spaced 10.6 Å/2 (or 5.3 Å). (**b**) Precipitate in the grain boundary. (**c**) shows edge dislocations in a precipitate. (**d**) Edge dislocation in high-resolution TEM image of (200) planes in fcc palladium

stress (or strain) field generated as shown in Fig. 8c, which can expand along a slip plane to produce a resolved shear stress and the production of a dislocation or multiple dislocations. In their work on observing dislocations using etch pits in LiF as illustrated in Fig. 6, Gilman and Johnston (1962) showed that even a dust particle falling on a single-crystal (100) LiF surface could produce sufficient, localized stress to create small dislocation loops produced at this surface source. With the addition of an applied stress, these dislocations loops became larger and were able to glide into the crystal interior. In this regard, it has been demonstrated even for metal single-crystal tensile specimens that as the stress is raised, the crystal strains

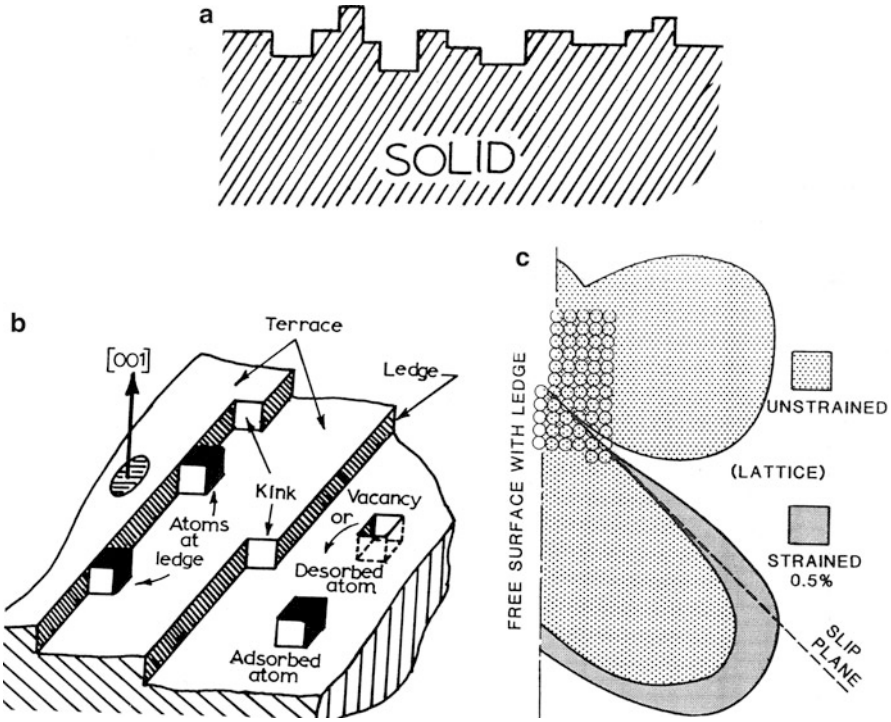


Fig. 8 Surface steps and ledge features. (**a**) Steps on solid crystal surface. (**b**) Surface ledge structure, including surface vacancy and adsorbed atom. (**c**) Computed strain incompatibility at single atom surface ledge

(deforms) as shown in Fig. 4a, but if the stress is released, and the tensile specimen surface uniformly machined away to a depth to which the surface-related dislocations have glided, reloading the specimen through the same stress regime will produce exactly the same stress–strain diagram. However, if the stress is continued, the stress–strain diagram will elevate somewhat exponentially and then become parabolic when the interior of the crystal becomes heavily dislocated and the dislocations interact and multiply to some critical level where fracture occurs. Figure 8 provides cartoons to illustrate these phenomena. Hirth (1963) has shown that the critical resolved shear stress for the production of dislocations in Al and Cu is roughly half at a surface step (as shown in Fig. 8c) in contrast to homogeneous nucleation of glissile dislocation loops discussed originally by F.C. Frank (1950).

Of course the operation of dislocation surface sources or their creation is not well understood and is related to the crystal orientation (and their activation implicit in Fig. 9f) and other features. As shown in Fig. 4b, the shear displacement forming a dislocation, and its glide to form a surface step opposite to the initial displacement, (shearing one crystal portion relative to the other by one displacement unit (or Burgers vector dimension)), differs from the prospect of continuous dislocation production from a surface ledge source. Because of the fundamental requirement of

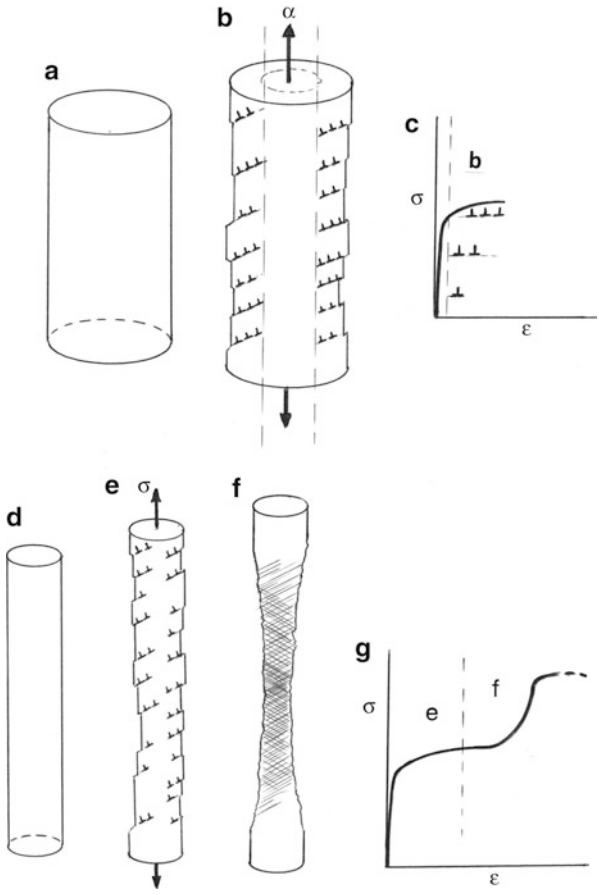


Fig. 9 Dislocation production at surface steps or ledges in a single crystal. (**a**) Perfect crystal cylinder. (**b**) After stress to produce surface-related dislocations. (**c**) Stress–strain diagram for (**b**). (**d**) Crystal cylinder machined to eliminate dislocations in (**b**). (**e**) Restressed cylinder as in (**b**). (**F**) Highly deformed cylinder showing necking and filling with dislocations. (**g**) Stress–strain diagram corresponding to (**f**). Note mass in (**d**–**f**) is constant

conservation of mass, a source cannot produce a continuous array of dislocations gliding into the crystal without moving mass from the surface, and ideally as shown in Fig. 4b, each interior dislocation requires a corresponding surface displacement relative to the glide (or slip) plane. As shown in Fig. 9, the necking which occurs during deformation provides this mass conservation in a complex way.

For the deformation of polycrystalline materials, internal surfaces (or interfaces) can provide the elastic incompatibilities required for dislocation production and even additional source mechanisms. Primary dislocation source mechanisms essentially start with the pinning of dislocations or dislocation segments as shown in Fig. 10a. At A (within the dotted circle) in Fig. 10a, the dislocation can act as a single-ended source as shown in Fig. 10b which has an immobile (or sessile) component (dashed) out of the slip plane and allows the dislocation to expand and spiral amount this sessile portion under an applied stress. Such spiral sources can be created when one end of a climbing edge dislocation is pinned (as at A in Fig. 10b), where it may be recalled that climb of an edge dislocation can be envisioned as the extra half plane (edge dislocation line) in

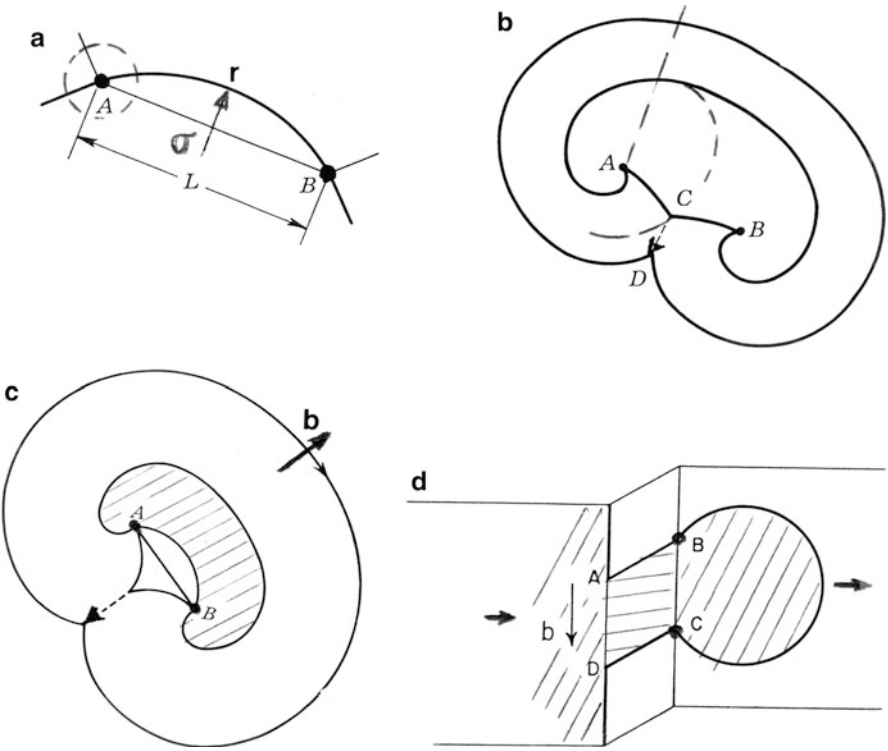


Fig. 10 Primary dislocation source mechanisms. (**a**) Single-ended (*A*) and double-ended (*A–B*) dislocation production. Segment length, *L*, experiencing stress σ bows to radius, *r*, optimized at $r = L/2$. (**b**) Single- (*A*) and double-ended (*A–B*) Nabarro–Herring source schematic. *Dotted* segment around *A* produces a spiral. (**c**) Frank–Read source in glide plane where dislocation segment (*L*) is pinned at *A* and *B* to produce continuous loops in the glide plane. (**d**) Double cross-slip Koehler source. Glide-plane segment is pinned at *B* and *C*

Fig. 5a recedes (or moves up) as a consequence of vacancy "condensation" along the dislocation (line). This usually occurs at elevated temperature and where large numbers of vacancies are present.

When a dislocation segment is pinned at two points (A and B) in Fig. 10a–c, the instability to activate this configuration as a dislocation source in the slip plane occurs ideally when the applied (or resolved) stress becomes:

$$\sigma \cong \frac{2Gb}{L} \tag{9}$$

where L is the length of the pinned segment, which bows under the applied stress to some critical stress where the maximum radius of curvature of the bowed section is $r = L/2$. As shown by the arrow in Fig. 10b from C to B, the bowed loop connects forming a continuous loop around A–B, and the process can continue under a constant, critical stress. Figure 10a, b characterizes the operation of Bardeen–Herring

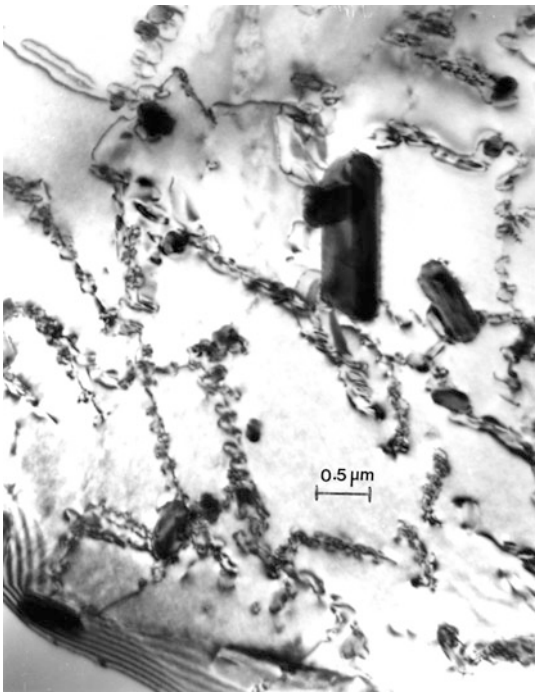


Fig. 11 Dislocation helices formed in a deformed Al-6061 alloy which was simultaneously heated and cooled, creating precipitates which pinned mobile (climbing) dislocations, creating helices

sources (Bardeen and Herring 1952), while Fig. 10c illustrates the Frank–Read (Frank and Read 1950a, b) source concept which is essentially the same in the context of dislocation production in the glide plane. Figure 10d shows a source proposed by J.S. Koehler (1952) which involves a double cross-slip mechanism with the dislocation segment pinned at C–B (in Fig. 10d) acting as the source. But Fig. 10d can also be associated with a ledge in an interface where, in contrast to cross-slip, dislocations gliding in the interface can provide both the ledge formation and its operation as a dislocation source in a glide plane.

While there are numerous observations of helical dislocations and dislocation helices produced by some form of single-ended source operation, as illustrated in Fig. 11, there are only a few recorded observations of Frank–Read source operation in polycrystalline materials, particularly stressed (or deformed) metals and alloys over the past 60 years. Figure 12 illustrates one of these few observations made by C.F. Bilsby in 1965 for an fcc Ni–Fe alloy age hardened and water quenched from 1,200 °C, by transmission electron microscopy. Dislocations were first observed by R.D. Heidenrich in 1949, but not recognized as dislocations in the TEM. P.B. Hirsch and colleagues identified the TEM images of dislocations a year later.

Although Frank–Read sources could be formed in deformed or stressed materials as illustrated in Fig. 11, the dearth of such observations in deformed or deforming engineering (polycrystalline) metals and alloys is an indication that dislocations are produced by other mechanisms other than Frank–Read source operation.

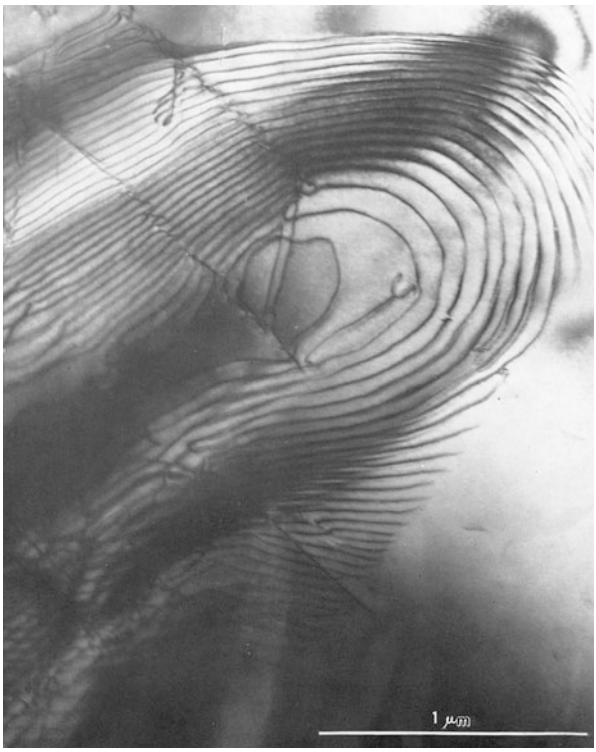


Fig. 12 Frank–Read dislocation source observed in age-hardened, Ni–Fe fcc alloy quenched from 1,200 °C. The source lies in the (111) (slip) plane which is in the plane of the page (TEM bright-field image courtesy of C.F. Bilsby)

In well-annealed polycrystalline metals and alloys, the only prominent strain instabilities can occur at interfaces (grain or phase boundaries), including of course the free surface (solid–vapor interface shown in Fig. 8). Figure 13 illustrates dislocation emission from sources in the grain boundaries of strained 304 stainless steel. It is apparent that as the engineering strain is increased (from 0.4 % to 1.5 % in Fig. 13), the source density in the boundary plane increases along with the average number of dislocations in the source emission profiles. Figure 14 illustrates these sources to be grain boundary ledges which can form by the glide of dislocations in the boundary plane (grain boundary dislocations: GBDs) and their interaction (Fig. 14a, b), as well as the heterogeneous shear of the boundary by dislocations piling up against the boundary plane (Fig. 14c, d). The geometric/crystallographic features of grain boundary ledge sources are illustrated in Fig. 14e, f. The formation of ledge sources illustrated in Fig. 14a, b could involve some form of the Koehler source shown in Fig. 9d. Figure 15 illustrates computed stress fields for a parallel edge ledge in a symmetric tilt (grain) boundary. The configuration in Fig. 15a promotes crack initiation which is illustrated in ▶ Fig. 23 of chapter "A Brief Introduction to Quantum Mechanics", while the ledge shear stress configuration in Fig. 15b promotes dislocation production. It can be noted in Fig. 15 that considerably more stress is required to initiate grain boundary (ledge-related) cracks (Fig. 16).

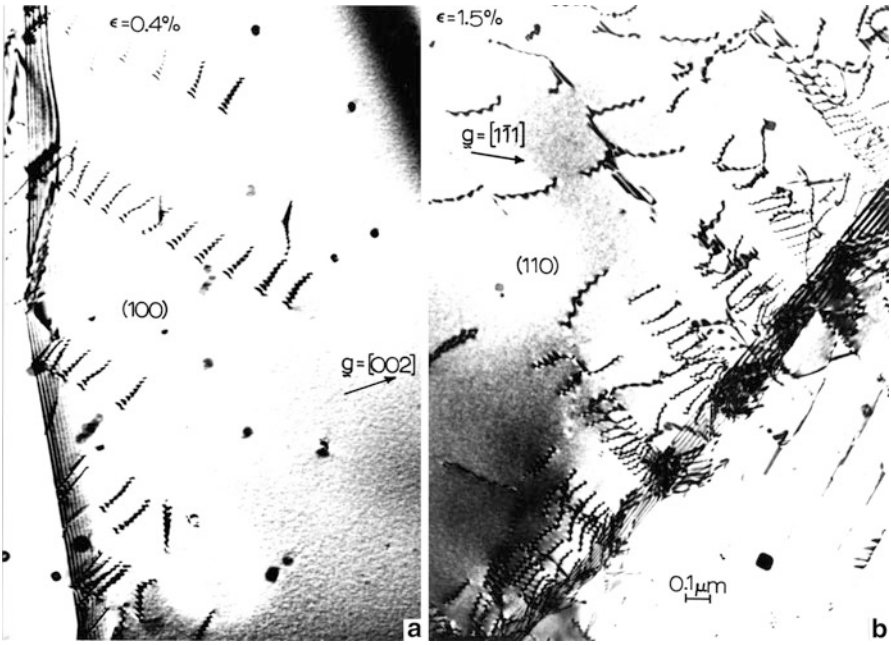


Fig. 13 Dislocation emission from grain boundary ledge sources in 304 stainless steel deformed in tension. TEM images at magnification shown in (**b**). (**a**) 0.4 % engineering strain. (**b**) 1.5 % engineering strain. The operating reflection for the image contrast is shown by *g*

J.C.M. Li (1963) was one of the first to recognize the importance of grain boundary dislocation ledge sources, and they have become the recognized and preferred mechanism for dislocation production in polycrystalline metals and alloys (Murr 1975a, 1975b; Murr and Wang 1982; Murr et al. 1970; Cass 1970; Bernstein et al. 1972; Marcinkowski 1972; Conrad et al. 1972). Murr and Wang (1982) performed a careful analysis of dislocation production in polycrystalline 304 stainless steel (a low-stacking-fault free energy alloy) and nickel (a high-stacking-fault free energy metal) at various stages of deformation in tension. Figure 17 shows the selection of strained specimens extracted along the stress–strain diagram for 304 stainless steel and observed in the TEM. Figure 18 illustrates typical, bright-field TEM images representing the corresponding engineering strains (Fig. 17) in percent, from ~0.05 % to 25 % strain. It can be noted, as illustrated in Fig. 13, that as the strain is increased, the ledge density (number/cm of boundary length) increases along with the number dislocations in the ledge source emission profiles: There is also a corresponding increase in the dislocation density in a region proximate to the grain boundary. These quantitative issues are summarized for the observations in 304 stainless steel (Fig. 18) in Fig. 19. It can be observed that the number of ledges/unit length of grain boundary (N/μm) or the active ledge source represented by dislocation profile frequency increases with increasing strain to values near 8/μm or ~0.8×10^5/cm. This is commensurate with maximum grain boundary

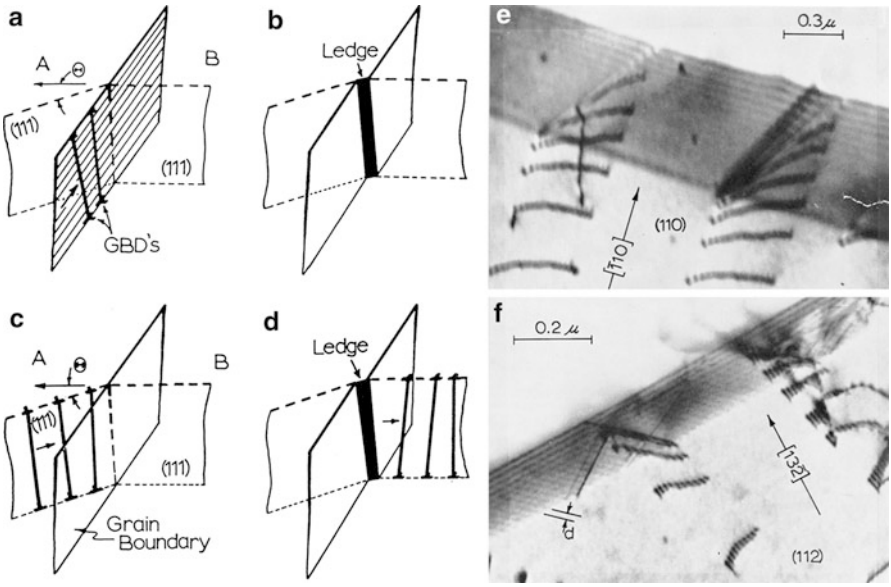


Fig. 14 Grain boundary ledge formation and dislocation emission from grain boundary ledge sources in 304 stainless steel. (**a**) and (**b**) show ledge formation by grain boundary dislocation (*GBD*) glide. (**c**) and (**d**) show ledge formation by heterogeneous shear of the boundary by dislocation pileup. (**e**) and (**f**) show TEM images of dislocation emission from grain boundary ledges ((**a**–**d**) are adapted from Murr (1975a))

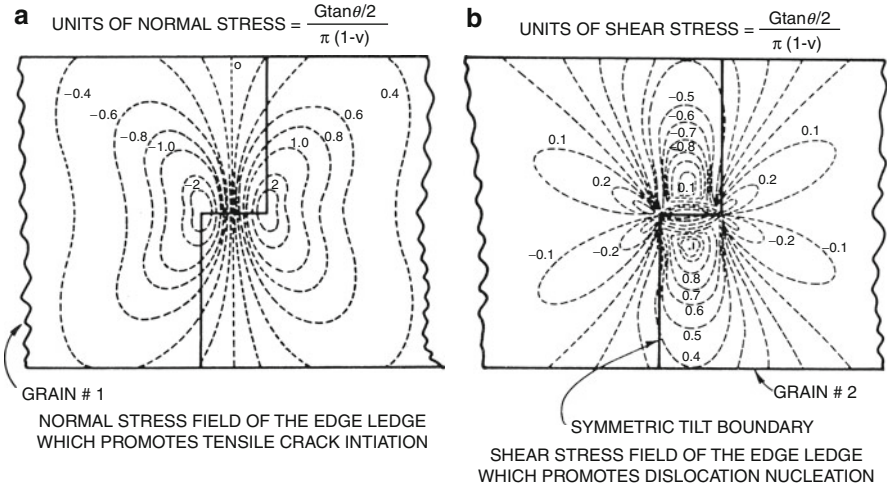


Fig. 15 Computed normal (**a**) and shear (**b**) stress fields for grain boundary ledge. (**a**) promotes crack formation, while (**b**) promotes dislocation production (Courtesy of Dr. E.S.P. Das)

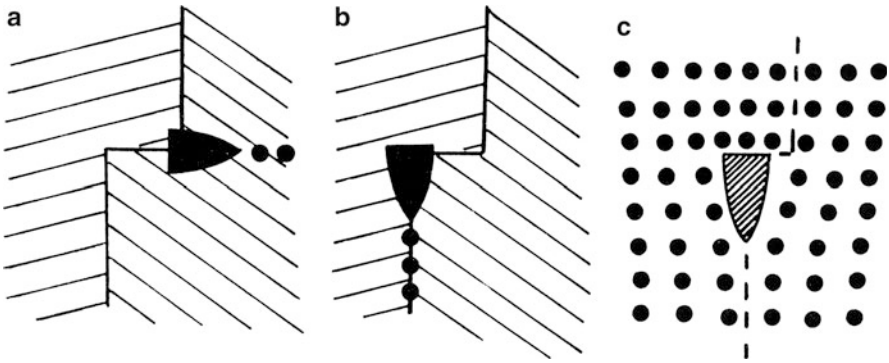


Fig. 16 Microcrack initiation at grain boundary ledges. (**a**) Transgranular microcrack. (**b**) Intergranular microcrack. (**c**) Classical edge dislocation coalescence mechanism of Stroh observed to resemble a grain boundary ledge. The wedge-shaped crack nucleus is shown *shaded* (Adapted from Murr (1975a))

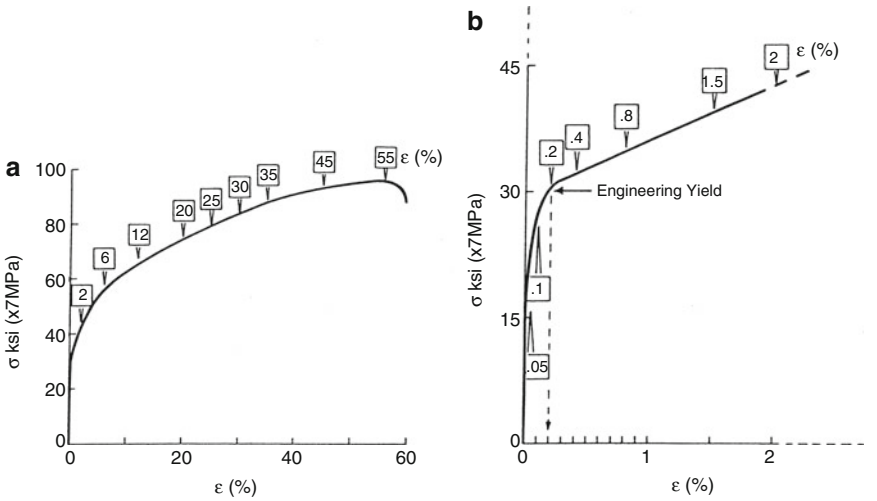


Fig. 17 Engineering stress–strain diagrams for 304 stainless steel sheet. (**a**) Complete strain range to fracture. "Boxes" indicate specific strains at which TEM samples were extracted for observation of microstructures. (**b**) Low-strain range expanded diagram (From Murr and Wang (1982))

ledge densities of ~0.7 × 10^5/cm measured by Bernstein et al. (1972) for strained iron and 4 × 10^5/cm for deformed titanium by Conrad et al. (1972).

Hirth and Lothe (1968) have also described so-called wedge dislocation sources whose specific mechanism of operation is largely unknown. Figure 20b illustrates this geometry of dislocation emission from a grain boundary in contrast to the ledge source as shown for comparison in Fig. 20a. Figure 21 compares two examples of wedge source emission observed directly in the high-voltage TEM for in situ straining of a thin 304 stainless steel foil. These source observations are more

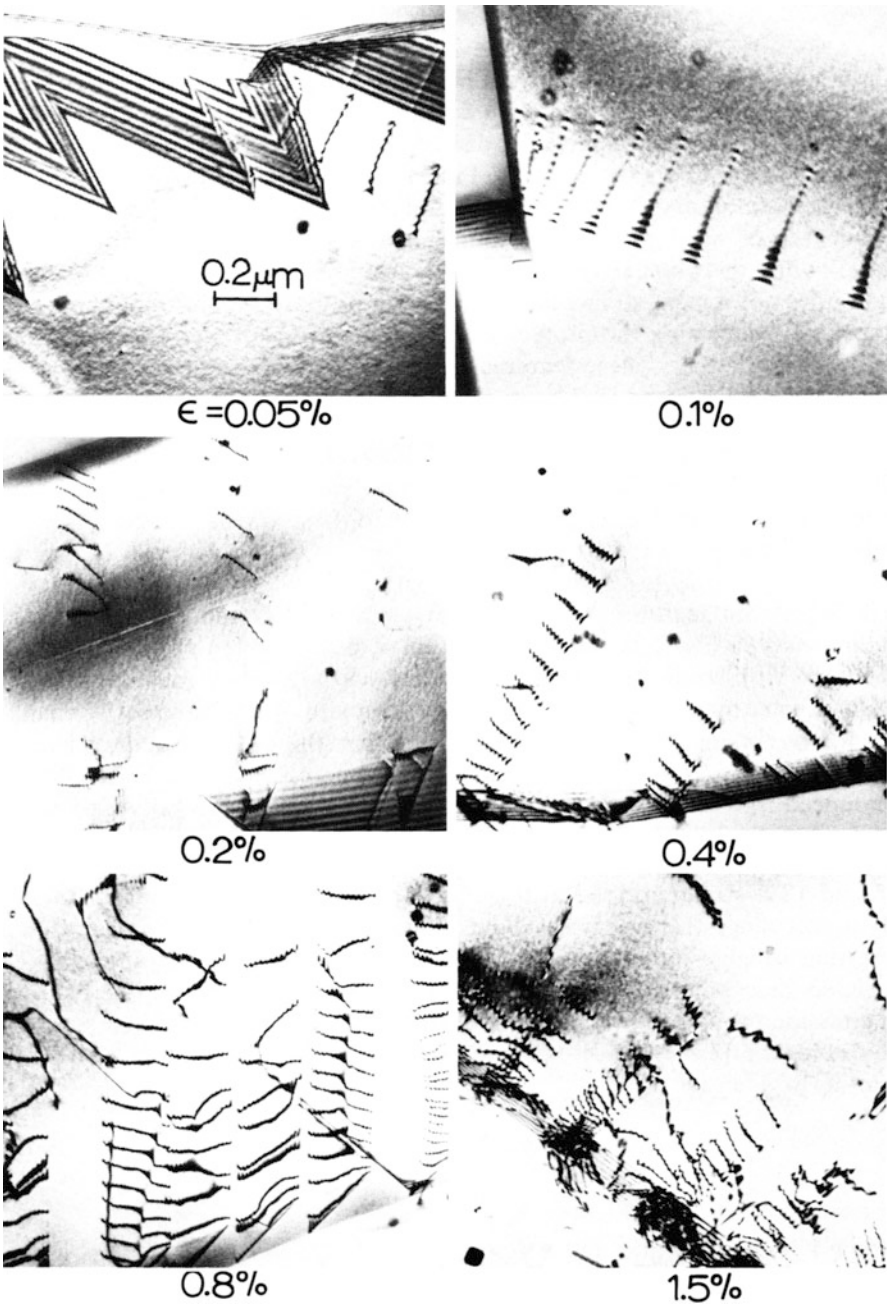


Fig. 18 (continued)

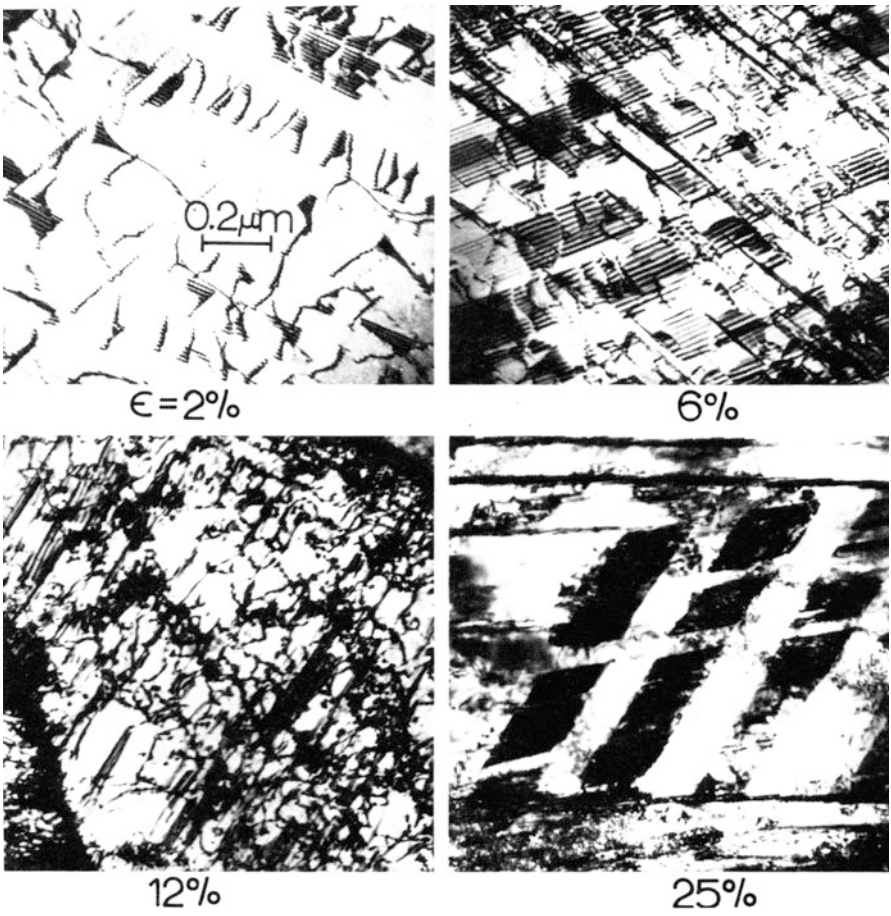


Fig. 18 TEM images showing grain boundary ledge emission of dislocations for deformed 304 stainless steel corresponding to strain points in ▶ Fig. 24 of chapter "A Brief Introduction to Quantum Mechanics". Magnifications are shown for $\epsilon = 0.05$ %. Grain surface orientations are (110) (From Murr and Wang (1982))

frequent than Frank–Read source operation, but certainly not as prevalent as ledge source production of dislocations.

In some respects, dislocation emission from ledge sources in the grain boundaries shown in Fig. 18, up to strains <1.5 %, is similar to the production of dislocations at free surfaces in single-crystal specimens shown schematically in Fig. 9b, e. The filling of grain interiors and dislocation interaction is observed at 2 % strain in Fig. 18, while at 6 % and 12 % strain, there is profuse reaction and interaction of dislocations, including profuse dissociation into stacking faults at 6 % strain (Fig. 18) and even α' martensite production (phase transformation) at 25 % strain in Fig. 18.

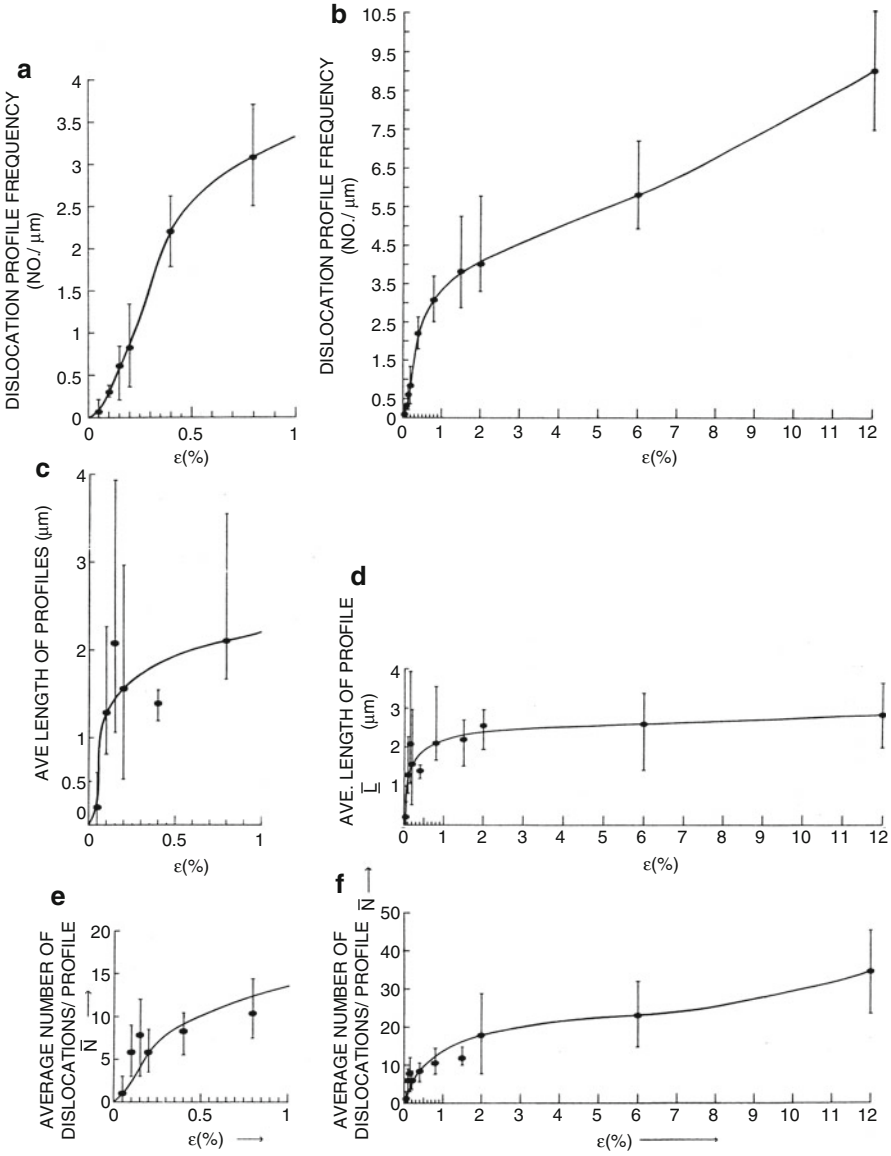


Fig. 19 Quantitative analysis of the dislocation emission profiles characterizing the lower-strain regime in uniaxially deformed 304 stainless steel. (**a**) Dislocation profile frequency versus strain in the low-strain range. (**b**) Dislocation profile frequency versus strain over a large strain range (to 12 % strain). (**c**) Average dislocation emission profile length versus strain in the low-strain range. (**d**) Average dislocation emission profile length versus strain in a larger strain range (to 12 % strain). (**e**) Average number of dislocations composing individual emission profiles versus strain in the low-strain range. (**f**) Average number of dislocations composing individual emission profiles versus strain in a larger strain range (to 12 % strain) (From Murr and Wang (1982))

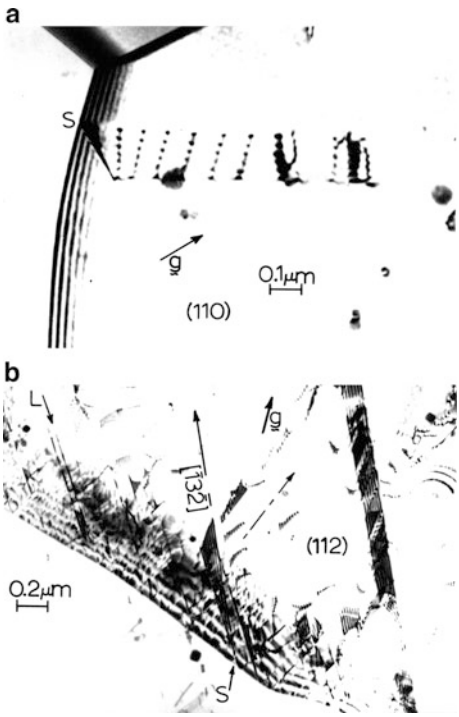


Fig. 20 Grain boundary ledge source (S) (**a**) and wedge source (S) (**b**) in 304 stainless steel

These features are also apparent in the progression of dislocation production and interaction corresponding to the stress–strain diagram for deformed stainless steel sheet by the penetration of a 0.22 caliber ballistic projectile as shown in Fig. 22. It is interesting to compare the hardness and strain values shown in Fig. 22 with the dislocation images at equivalent strains in Fig. 18, even though the bullet hole example in Fig. 22 represents a strain rate nearly one million times greater than the tensile deformation represented in Fig. 18.

It can be observed from Fig. 19 that dislocation density as expressed by dislocation profile lengths and numbers of such profiles as related to ledge sources in the grain boundary is related to engineering strain in some form:

$$\varrho = \varrho_0 + A\epsilon^m \tag{10}$$

where ϱ is the dislocation density in number/cm^2, ϱ_0 is the incipient or initial dislocation content, and A is some dislocation multiplication constant which might include both free surface ledge density and the ledge density in the grain boundaries. The exponent, m, in Eq. 19, as observed in Fig. 19, can be approximated by m ≅ 1/2 for A ≅ 3. However, the associated stress becomes

$$\sigma = \sigma_0 + \mathrm{KGb}\sqrt{\varrho}, \tag{11}$$

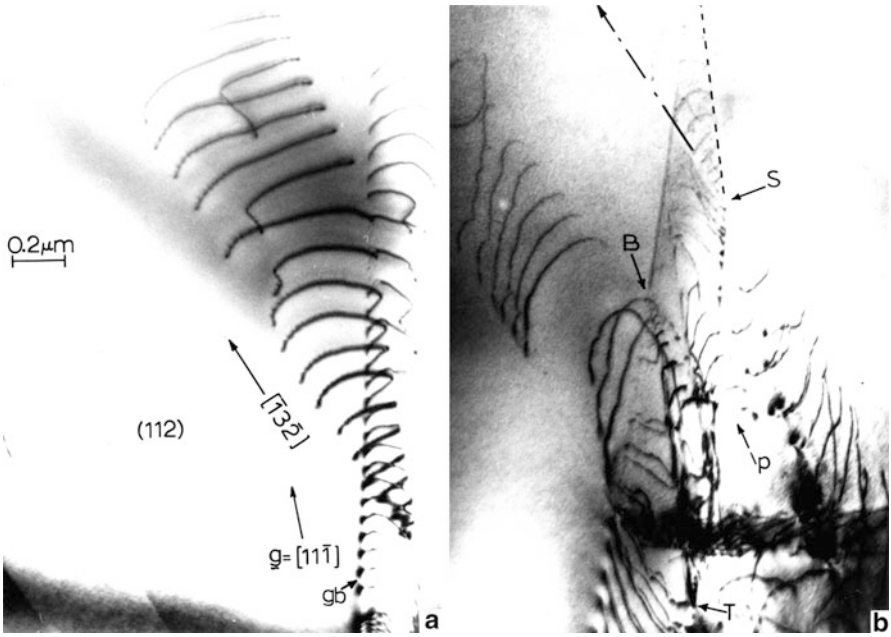


Fig. 21 Operation of wedge-type dislocation sources while straining thin 304 stainless steel foil in a high-voltage (1 MV) electron microscope. (**a**) Grain boundary dislocations activating a wedge source. (**b**) Wedge source operating from grain boundary at S and ledge source operating along T–B indicating *top* (T) and *bottom* (B) of thin foil under tensile stress. P denotes dislocations moving into the grain boundary from pileup

where

$$\sqrt{\varrho} = A\sqrt{\epsilon}. \tag{12}$$

Equation 12 may also apply to single crystals, but for a polycrystalline metal or alloy (Conrad et al. 1972):

$$\sqrt{\varrho} = A\sqrt{\epsilon} + B/\sqrt{D}, \tag{13}$$

where the first term applies mainly to free surface dislocation generation while the second term relates to the grain boundary. However, we find, or squaring Eq. 13 (Conrad et al. 1972),

$$\varrho = \underset{\left(\begin{matrix}\text{Single}\\ \text{Crystal}\end{matrix}\right)}{A^2\epsilon} + \underset{\begin{matrix}\text{Strain + grain size}\\ \text{multiplication}\end{matrix}}{2AB\sqrt{E/D}} + \underset{\left(\begin{matrix}\text{Grain size}\\ \text{effect}\end{matrix}\right)}{B^2/\sqrt{D}}, \tag{14}$$

where D is the grain size and A^2, AB, and B^2 are dislocation multiplication constants, and the middle term will include multiplication of dislocations both at grain boundaries and by dislocation interactions or reactions within the grains, especially at higher engineering strains; $\epsilon > 6$ % in Fig. 18.

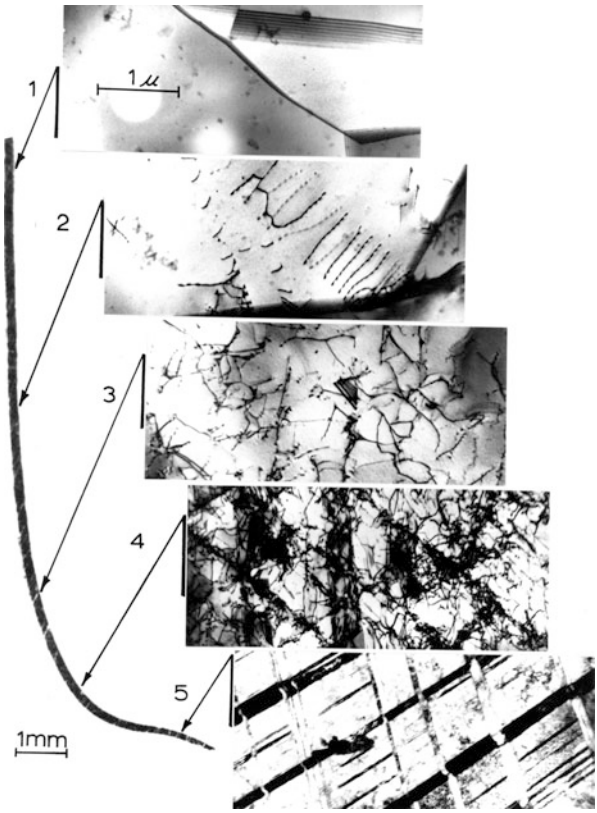


Fig. 22 TEM images corresponding to straining of bullet hole petal observed edge-on following 0.22 caliber projectile penetration through thin 304 stainless steel sheet. (1) $\epsilon = 0$ %, HV = 1.4 GPa; (2) $\epsilon = 2$ %, HV = 2.0 GPa; (3) $\epsilon = 6$ %, HV = 2.2 GPa; (4) $\epsilon = 12$ %, HV = 2.8 GPa; (5) $\epsilon = 25$ %, HV = 3.8 GPa (Adapted from Murr and Foltz (1970)). Note that as a "rule of thumb," $\sigma \cong$ HV/3

Considering the quantification of dislocation multiplication implicit in the ledge formation developed by Li (1963) and Li and Chou (1970), we can let

$$B^2 = KGb\left(\frac{m}{\pi}\right)^{1/2}, \tag{15}$$

where m is the ledge density in number of ledges/cm of grain boundary which, according to Fig. 19, is also strain dependent. Dislocation production will also depend on the so-called *slip length* which decreases with decreasing grain size, and this phenomenon is somewhat implicit in the middle term of Eq. 14. The constant K in Eq. 15 can be ~3.

Equations 11, 12, 13, 14, and 15 provide some semiquantitative implications for grain size, ledge density, and the prospects for hardening as dislocation density increases with strain, and additional multiplication of dislocations occurs through reaction and interaction in the grain interiors. These phenomena will also be influenced by stacking-fault free energy in fcc metals and alloys as this sharply influences slip and cross-slip, and corresponding cross-slip-related dislocation source production implicit in Fig. 10d.

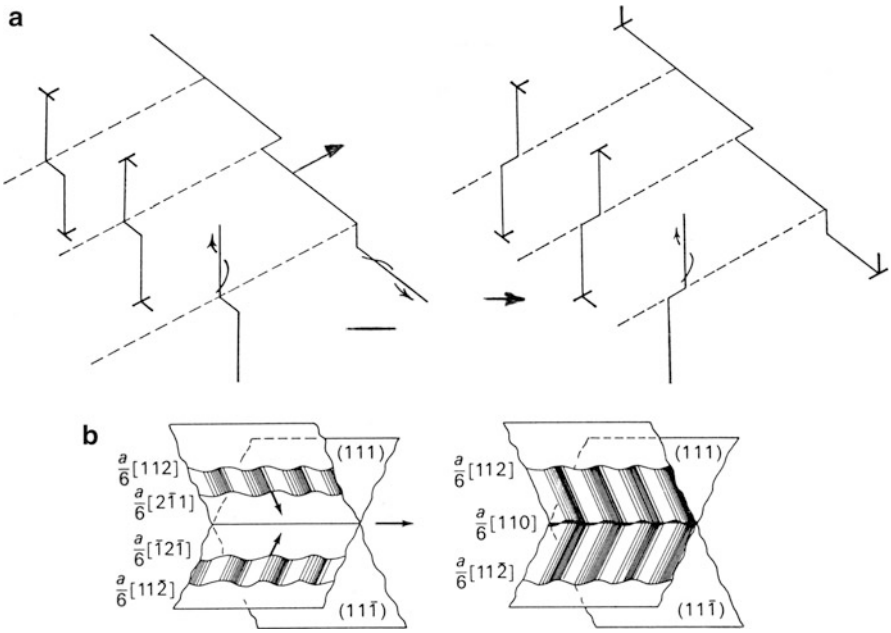


Fig. 23 Dislocation intersections (After Cottrell (1953)) forming jogs or steps (**a**). Partial dislocations connected by a stacking-fault region intersecting to form a sessile lock (Cottrell-Lomer lock which cannot glide on the {111} slip planes)(**b**)

While the dislocation ledge sources illustrated for low strain in stainless steel in Fig. 18 show dislocation emission on a single slip system, as the strain is increased, dislocations are emitted from other slip-oriented ledges in opposing grain boundaries. This raises the propensity for dislocation interactions as shown in Fig. 23 (Cottrell 1953). These can form various types of jogs which can move conservatively or nonconservatively, even forming dislocation dipoles at pinning points and pinching them off to form dislocation loops.

Dislocations, partially connected by a region of stacking fault as shown in Fig. 23b, can also form a sessile lock which can act as a barrier to dislocations and create small pileups in the grain interior. Many of these features have been treated in detail by Hirth and Lothe (1968). In the case of fcc metals and especially low-stacking-fault free energy alloys, it is probably useful to review the nature of dislocation reactions and dissociations on {111} planes, especially in the context of energy minimization in this context; Figure 24 shows an unfolded net for the Thompson tetrahedron formed from the four unique {111} planes in fcc: (111) (d), $(1\overline{1}\overline{1})$ (b), $(\overline{1}1\overline{1})$ (a), and $(\overline{1}\overline{1}1)$ (c). The total dislocations with Burgers vectors of the form a $<$ 100$>$/2 form the edges (**AB**, etc.) of the tetrahedron, while the associated (or dissociated) partial dislocations with Burgers vectors of the form a $<$ 112$>$/6 are shown as **A**$\boldsymbol{\delta}$, **B**$\boldsymbol{\delta}$, and **C**$\boldsymbol{\delta}$ in the (111) plane, where δ, γ, β, and α designate the midpoints of the corresponding {111} planes. Only the vector indices

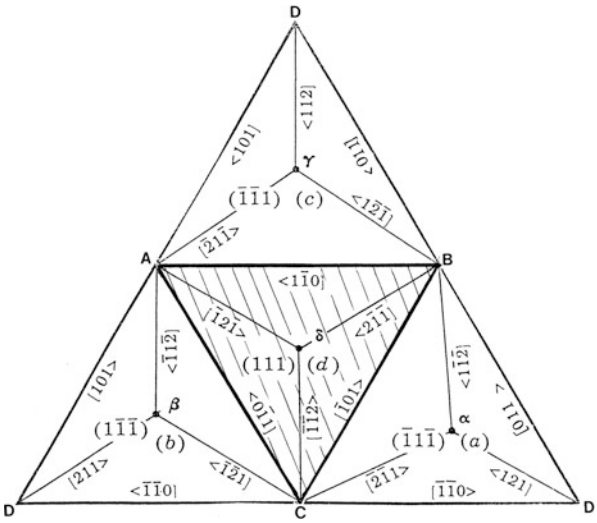


Fig. 24 Unfolded {111} Thompson tetrahedron showing total and partial dislocation Burgers vectors for dislocations glissile on {111} planes ((111) plane is *shaded*). Net can be folded to D (vertex). Directions are indicated <or>. Changing direction changes sign of indices (adapted from Hirth and Lothe (1968))

and directions (<, >) are shown in Fig. 24. Note that in adding or subtracting any Burgers vector, the directions must be consistent or accounted for in the signs of the Miller indices. For example, the total dislocation Burgers vectors at the (111) plane edges can glide in (111) and in the other planes represented by the corresponding edges. If we have dislocation Burgers vectors along **AC** and **CB** which glide in the (111) plane and become parallel to one other, they can react in the following ways:

$$\mathrm{a}\left[01\bar{1}\right]/2 + \mathrm{a}\left[\bar{1}01\right]/2 = \mathrm{a}\left[\bar{1}10\right]/2, \tag{16}$$

$$\mathrm{a}\left[01\bar{1}\right]/2 - \mathrm{a}\left[\bar{1}0\bar{1}\right]/2 = \mathrm{a}\left[\bar{1}0\bar{2}\right]/2, \tag{17}$$

In the first reaction, two total dislocations (with Burgers vectors shown) combine to form another total dislocation whose energy is $a^{2}/_{2}$ (according to Eq. 16). Correspondingly, in Eq. 17, the dislocation formed would have an energy of $3a^{2}/_{2}$, and this would be energetically unfavorable. Conceptually, the addition of dislocations with Burgers vectors **AC** and **CB** in Fig. 24 equals **AB**. Consequently these can all mutually intersect at a node (or triple node) on the (111) plane center (at d).

Each total dislocation with Burgers vectors at the {111} plane edges in Fig. 24 can dissociate into partial dislocations having Burgers vectors of the form a < 112>/6:

$$\left.\begin{array}{l} a\left[1\bar{1}0\right]/2 \rightarrow a\left[1\bar{2}1\right]/6 + a\left[2\bar{1}\bar{1}\right]/6 \\ a\left[0\bar{1}1\right]/2 \rightarrow a\left[\bar{1}\bar{1}2\right]/6 + a\left[1\bar{2}1\right]/6 \\ a\left[\bar{1}01\right]/2 \rightarrow a\left[\bar{1}\bar{1}2\right]/6 + a\left[\bar{2}11\right]/6 \end{array}\right\} \tag{18}$$

or energetically

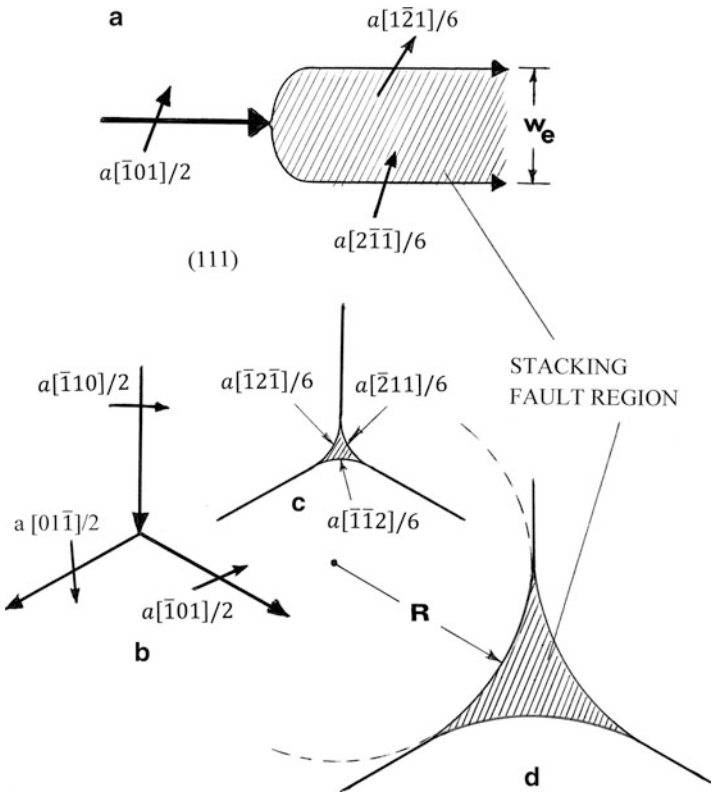


Fig. 25 Dislocation dissociation into lower-energy partials on fcc (111) (**a**). Dissociation (extension) of a dislocation node (**b**) forming a stacking-fault region (*shaded* in (**c**) and (**d**)). The partial dislocations in (**a**) are separated by a stacking fault (*shaded*). The equilibrium separation is w_e

$$\left(a/2\right) \rightarrow \left(a/6\right) + \left(a/6\right) = \left(a/3\right) \tag{19}$$

It can be noted in Eq. 18 that each total dislocation in the (111) plane shares a partial dislocation, and in the case of a dislocation triple node, the node can correspondingly dissociate or extend, with total dislocations sharing partials. These reactions are illustrated schematically in Fig. 25.

Both the equilibrium separation (w_e) of partial dislocations, $\mathbf{b_p}$, in Fig. 25a and the radius of curvature (R) for an extended node as shown in Fig. 25d are related to the stacking-fault free energy γ_{SF}:

$$\gamma_{SF} = \frac{G\left|\mathbf{b}_p\right|}{8\pi w_e}\left[\frac{2-\upsilon}{1-\upsilon}\left(1-\frac{2\upsilon\cos 2\alpha}{2-\upsilon}\right)\right] \tag{20}$$

and

$$\gamma_{SF} = G\left|\mathbf{b}_p\right|/2R, \tag{21}$$

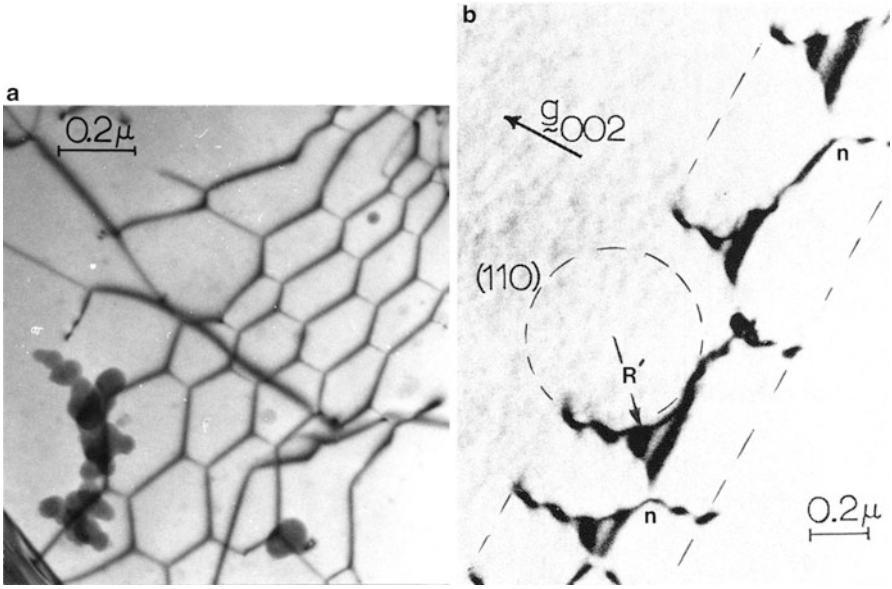


Fig. 26 Dislocation nodes. (**a**) Dislocation network of nodes in beryllium. Network is on inclined (001) plane. (**b**) Extended dislocation node network on inclined (111) plane in 304 stainless steel. Nodes at "*n*" are not extended, and connecting dislocation is invisible. The surface orientation is (110), and the operating reflection in the TEM image is [002]

where α in Eq. 20 is the partial dislocation character angle; $\alpha = 0$ for a screw dislocation at 90° for an edge dislocation. It can be noted in Figs. 13a and 14f that dislocations emitted from grain boundary sources are dissociated, but it is difficult to know if the separation represents w_e in Fig. 25. Dislocation nodes exhibit extension (or dissociation) at 0.2 % strain in Fig. 18 as well, while extended, planar stacking faults are observed at a strain of 6 % in Fig. 18. We will discuss stacking-fault energies and stacking-fault phenomena more generally in chapter "► Planar Defects: Crystal Interfaces" which deals with planar defects.

It might be noted in Fig. 26 that dislocation nodes on inclined slip (or glide) planes represent a distorted geometry in the projected image in the TEM which must be corrected if accurate measurements of R (and correspondingly γ_{SF}) in Eq. 21 and Fig. 25d are to be made (Murr 1975a). In Fig. 26a, dislocation nodes form a hexagonal network on the (001) hcp Be basal glide plane which is distorted in the projection of this inclined (001) plane in the TEM image. Figure 26b shows a similar node arrangement on the (111) plane in 304 stainless steel where the projected, extended-node radius, R□, must be corrected for the projection distortion. It can be noted in Fig. 26a that the nodes are not recognizably extended while in Fig. 26b nodes indicated by (n) are also not extended, and the dislocation from the extended (stacking-fault) node is invisible, indicating that $\mathbf{g}, \mathbf{b} = 0$, and therefore since $\mathbf{g} = [002]$ in Fig. 26b, $\mathbf{b}$ is of the form [110] [73].

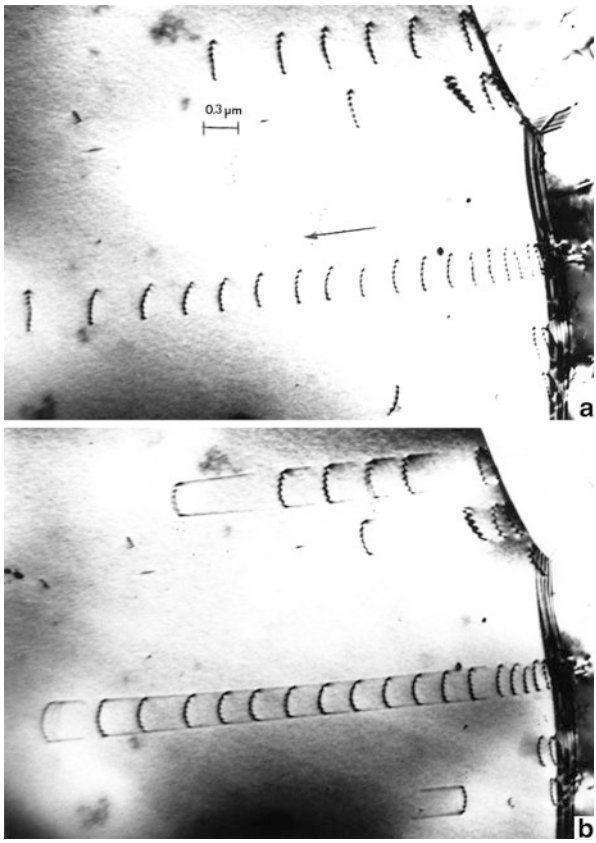


Fig. 27 Dislocation emission profiles at grain boundary sources. (**a**) Initial image. (**b**) Dislocation motion after dwelling beam on the area to create complex thermal stress profile. Traces of $(1\bar{1}1)$ planes corresponding to $[32\bar{1}]$ direction shown by *arrow* arise from surface contamination. The surface orientation is (103)

The strengthening or work hardening of metals and alloys is often related to dislocation pileups at grain boundaries or other obstacles such as the sessile dislocation lock in Fig. 23b. However, extensive pileups are not really observed in high-stacking-fault free energy fcc metals such as nickel and aluminum where dislocations easily cross-slip in contrast to 304 stainless steel in Fig. 13 or 18, where pileups are traditionally mistaken for dislocation emission profiles. Figure 27 shows long dislocation emission profiles in 304 stainless steel which, perhaps in Fig. 27a, at least, appear to be identical to pileups. Borrowing from dislocation pileup formation, the total force on the ith dislocation is given by (Weertman and Weertman, 1992)

$$F_i = \frac{Gb^2}{2\pi(1-\upsilon)} \sum_{\substack{j=0 \\ j=1}}^{n} \left(\frac{1}{x_i - x_j}\right) - \sigma b = 0, \tag{22}$$

where υ is Poisson's ratio and n represents the number of dislocations in the pileup. Correspondingly, the stress on the leading dislocation at the head of the pileup is given by

$$\sigma^* = n\sigma \tag{23}$$

where σ is the applied stress. But if the pileup is an emission profile, the stress, σ^*, in Eq. 23 is the operating stress at the ledge source. In Fig. 27b, the electron beam dwelling in the image area causes an irregular (thermal) stress to act on the dislocations in the profiles, and they bow and move, leaving a trace at the surface of the thin film due to the formation of a surface contamination film produced by the beam. While dislocations in the larger profiles in Fig. 26b move to the left, the single dislocation near the bottom right is observed to move to the right. The acting stress for dislocation pileups is expressed approximately by

$$\sigma \cong \frac{nGb}{\pi l}, \tag{24}$$

where n is the number of dislocations in a pileup of length, l. If this is applied to the dislocation emission profiles as determined quantitatively in Fig. 19c, d, n/l is observed to increase from ~5 μm^{-1} at 1 % strain (ϵ) to ~12 μm^{-1} at 12 % strain and a corresponding increase in stress (σ) as illustrated in Fig. 17.

The dissociated dislocations in emission profiles shown in Figs. 13 and 18 and the linear arrays of dislocations illustrated in Fig. 27 are characteristics of low-stacking-fault free energy (low γ_{SF}) alloys, 304 stainless steel in these examples: $\gamma_{SF} \cong 21$ mJ/m^2 (Murr 1975a). As shown in Eq. 19, dissociation of total dislocations in fcc metals and alloys is energetically motivated, but if the stacking-fault free energy is high, this must be accounted for in the complete energy consideration in Eq. 19. For example, extended stacking-fault regions in low-γ_{SF} alloys such as stainless steel mandate that slip occurs on the same slip plane, forming linear dislocation arrays as shown in Figs. 13, 18, and 27. However, the inability of dislocations to measurably dissociate for high-stacking-fault free energy metals such as Cu($\gamma_{SF} \cong 80$mJ/m^2), Ni($\gamma_{SF} \cong 130$ mJ/m^2), and Al($\gamma_{SF} \cong 180$ mJ/m^2) (Murr 1975a), for example, favors cross-slip as represented in Fig. 28. The emission of dislocations from a grain boundary (ledge) source in Al in Fig. 28b illustrates profuse cross-slip (sc), and when such high-γ_{SF} metals are deformed, they do not normally produce recognizable emission profiles such as those shown in Fig. 18, nor do they produce any recognizable dislocation pileups. In fact, dislocations emitted from grain boundary ledge sources in high-γ_{SF} metals quickly interact by cross-slip near the boundary, forming dislocation angles which gradually extend into dislocation cell-like arrays as illustrated for Cu in Fig. 29a, b for coherent twin boundaries. For comparison, Fig. 19c shows dislocations emitted and possibly piling up at coherent twin boundaries in 304 stainless steel. Figure 29a–c can be compared directly with the sketch for high-γ_{SF} and low-γ_{SF} metals and alloys depicted in Fig. 28a. Murr and Wang (1982) have also illustrated these features for tensile deformation of Ni.

Figures 30 and 31 illustrate the progression of dislocation cell formation implicit in Fig. 29a, b for Cu and Ni, respectively. Figure 31 shows the development of dislocation cells and cell wall structure in Ni, in contrast to deformed Cu in Fig. 30.

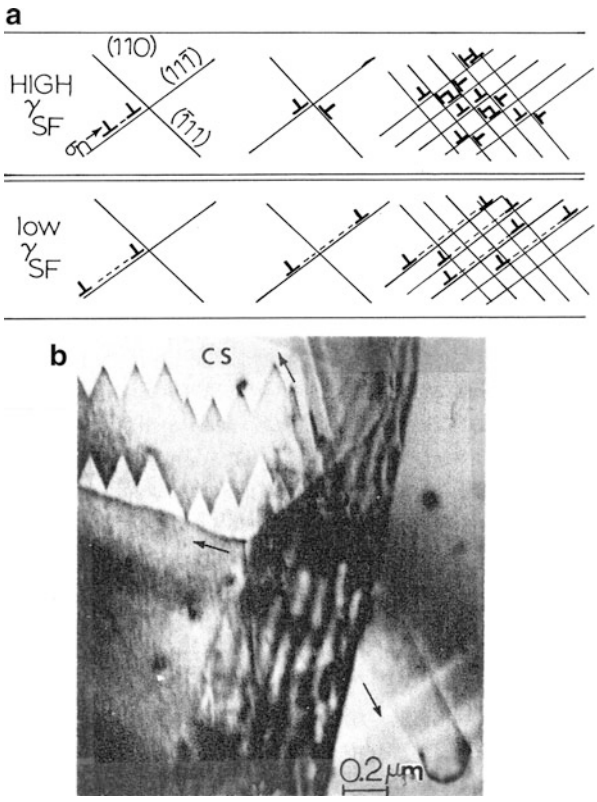


Fig. 28 Stacking-fault free energy (γ_{SF}) influence on dislocation dissociation and cross-slip. (**a**) High-γ_{SF} promoting dislocation cross-slip and interaction in contrast to low-γ_{SF} promoting linear dislocation arrays. (**b**) Dislocation emission and cross-slip from a grain boundary ledge source in aluminum (*cs*). The arrays show dislocations on multiple slip planes

The insert in Fig. 31 shows only remnants of cell structure in heavily deformed Ni. It is apparent from the dislocation densities and the reduction of the dislocation cell size, d, with deformation (stress), that there is a relationship between dislocation density and cell size, as well as cell size and stress. Figure 32 illustrates these features for plane-wave shock loading of Cu and Ni.

In perusing the examples of dislocation structures in stainless steel in Figs. 13 and 18 and in Cu and Ni in Figs. 30 and 31, it is apparent that for fcc metals and alloys, the stacking-fault free energy has a notable effect on the deformation–slip mechanism and the residual or associated deformation-related dislocation structures or substructures. This is also true for other crystal structures, particularly where the slip systems are different and the energetics of dislocation dissociation are also different. It can be observed in the dislocation network on the inclined (001) plane in hcp Be in Fig. 26a that there is some small dissociation of the dislocations but not to the extent represented in ▶ Fig. 33b of chapter "A Brief Introduction to Quantum Mechanics" for fcc stainless steel. While the slip plane in hcp metals is preferably (001), which is the close-packed plane as in a single {111} plane for fcc, there are only three associated slip directions or Burgers vectors in contrast to four {111} slip planes and three slip directions or Burgers vectors for each plane as illustrated in the Thompson tetrahedron of Fig. 24.

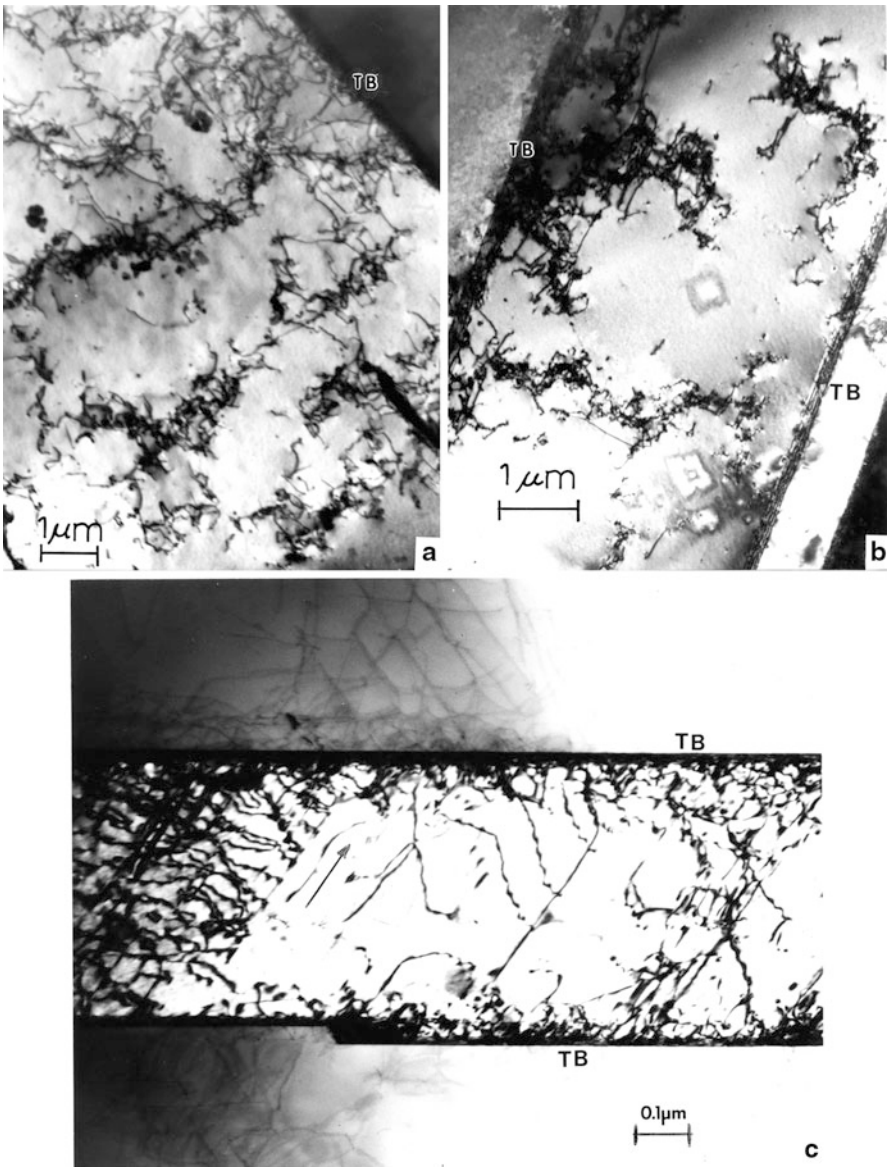


Fig. 29 Dislocation emission at ledges in coherent twin boundaries (*TB*) (coincident with {111} planes) in Cu (**a** and **b**), in stainless steel (**c**). The grain surface orientation in (**a**) and (**b**) is (100) and the twin boundaries are coincident with [011]. The grain surface in (**c**) is (110), and the twin boundaries (*TB*) are coincident with $[\bar{1}12]$. The *arrow* in (**c**) is the $[\bar{1}10]$ direction

Consequently, there are 12 slip systems in fcc. In contrast, bcc metals and alloys can slip on {110}, {112}, and {123} with corresponding $<111>$ slip directions. A comparison of these dislocation–slip features for fcc, bcc, and hcp materials is illustrated in Table 1.

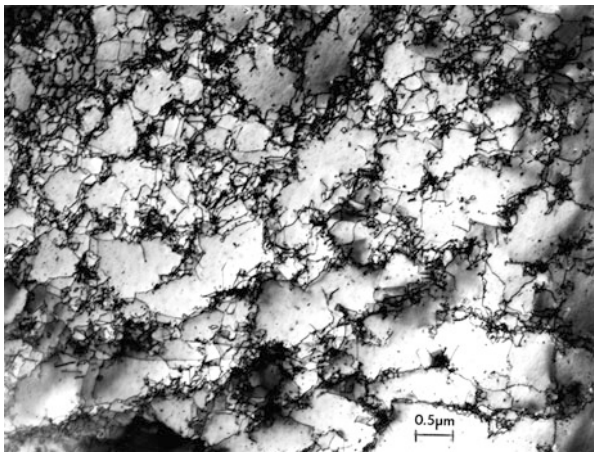


Fig. 30 Dislocation cell structure forming in the grain interior of deformed Cu

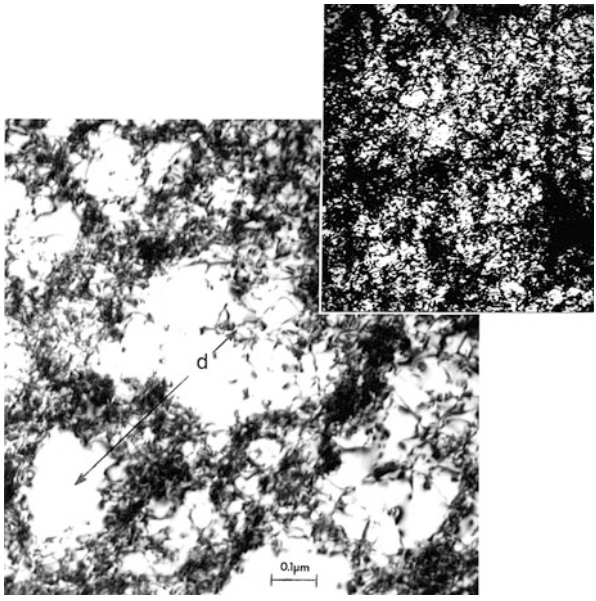


Fig. 31 Plane-wave shock deformed Ni. The main TEM image is Ni shock loaded at 15 GPa (150 kbars), while the insert represents Ni shocked at 30 GPa (300 kbars). The magnification of the insert is the same shown in the main image. The cell size convention is shown at *d*

While the primary slip planes in bcc are {110}, twinning occurs on {112}, and {123} slip can also occur at elevated temperature deformation. This slip–twinning variance in bcc in contrast to only {111} slip–twinning behavior in fcc assures recognizably different deformation microstructures in bcc materials in contrast to fcc materials in general. These will be explored in more detail in chapter "► Planar Defects: Crystal Interfaces" in dealing with planar defects. Figure 33 shows dislocation substructure in bcc Ta which appears similar to the dislocation substructure observed in fcc Cu in Fig. 29a, b.

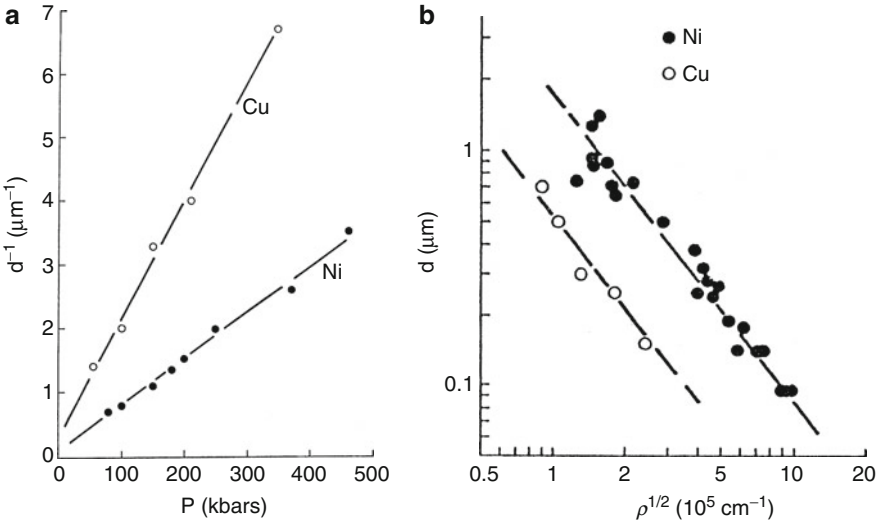


Fig. 32 (**a**) Reciprocal cell size (d^{-1}) versus plane-shock pressure (stress) (P) for Cu and Ni. (**b**) Dislocation cell size versus square root of dislocation density for plane-wave shock-loaded Cu and Ni

Table 1 Dislocation slip systems: fcc, bcc, and hcp crystal structures

Crystal structure	Slip planes	Slip directions	Number of slip systems
FCC	{111}	<110>	12
BCC	{110}	<111>	12
	{112}	<111>	12
	{123}	<111>	24
HCP	(001)	<110>	3

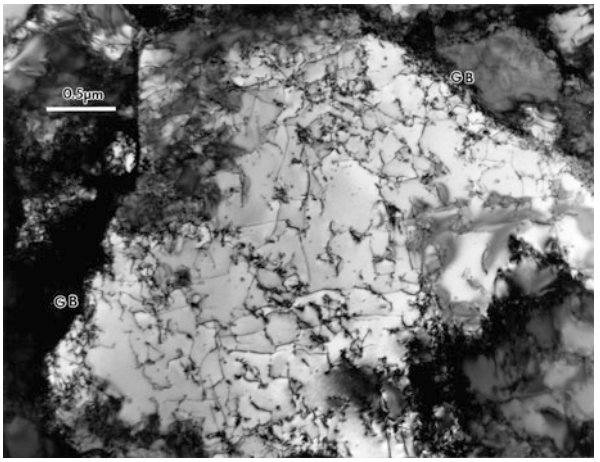


Fig. 33 Dislocation substructure in bcc Ta. Note dense dislocation arrays at the grain boundaries marked *GB* and presumed to be emitted from grain boundary ledge sources. The surface orientation is (001). Bright-field TEM image at 200 kV accelerating potential

References

Banhart F, Ajayan PM (1996) Carbon onions as nanoscopic pressure cells for diamond formation. Nature 382:433–435

Bardeen J, Herring C (1952) Imperfections in nearly perfect crystals. Wiley, New York, p 261

Bernstein IM, Rath BB, Thomas LE (1972) Grain boundary dislocations. In: Thomas G, Fulrath RM, Fisher RM (eds) Electron microscopy and structure of materials. University of California Press, Berkeley, pp 75–76

Bohsung J, Trebin HR (1987) Disclinations in quasicrystals. Phys Rev Lett 58(21):2277–2280

Cass TR (1970) The science, technology and application of titanium. Pengomn Press, New York, p 459

Conrad H, Okasaki K, Gadgil V, Jon M (1972) Dislocation structure and the strength of titanium. In: Thomas G, Fulrath RM, Fisher RM (eds) Electron microscopy and structure of materials. University of California Press, Berkeley, pp 438–469

Cottrell AH (1953) Dislocations and plastic flow in crystals. Clarendon, Oxford, UK

Frank FC (1950) Symposium on plastic deformation of crystalline solids. Carnegie Institute of Technology, Pittsburgh, p 89

Frank FC (1958) On the theory of liquid crystals. Discuss Faraday Soc 25:19–25

Frank FC, Read WT (1950a) Symposium on plastic deformation of crystalline solids. Carnegie Institute of Technology, Pittsburgh, p 44

Frank FC, Read WT (1950b) Multiplication processes for slow moving dislocations. Phys Rev 79:722–730

Gilman JJ (1969) Micromechanics of flow in solids. McGraw-Hill, New York

Gilman JJ, Johnston WG (1962) Dislocations in lithium fluoride crystals. In: Seitz F, Turnbull D (eds) Solid state physics, vol 13. Academic, New York, pp 148–222

Gilman JJ, Johnston WG, Sears GW (1958) Dislocation etch pit formation in lithium fluoride. J Appl Phys 29(5):747–754

Hirth JP (1963) Relation between structure and strength in metals and alloys. H.M. Stationery Office, London, p 218

Hirth JP, Lothe J (1968) Theory of dislocations. McGraw-Hill, New York

Kim P (2010) Across the border. Nat Mater 9:792–793

Koehler JS (1952) The nature of work hardening. Phys Rev 86:52–58

Leung PW, Henley CC, Chester GV (1989) Decagonal order in a two-dimensional Leonard-Jones System. Phys Rev B 39(1):446–459

Li JCM (1963) Petch relation and grain boundary sources. Trans TMS-AIME 227:239–247

Li JCM, Chou YT (1970) The role of dislocations in the flow stress grain size relationships. Metall Trans 1:1145–1150

Marcinkowski MJ (1972) Dislocation behavior and contrast effects associated with grain boundaries and related internal boundaries. In: Thomas G, Fulrath RM, Fisher RM (eds) Electron microscopy and structure of materials. University of California Press, Berkeley pp 383–416

Murr LE (1975a) Interfacial phenomena in metals and alloys. Addition–Wesley, Reading

Murr LE (1975b) Some observation of grain boundary ledges and ledges as dislocation sources in metals and alloys. Metall Trans A 6A:505–513

Murr LE, Foltz JV (1970) Anatomy of a bullet hole. J Mater Sci 5:63–72

Murr LE, Wang S-H (1982) Comparison of microstructural evolution associated with the stress–strain diagram for nickel and 304 stainless steel: an electron microscope study of microyielding and plastic flow. Res Mech 4:237–274

Murr LE, Horylev RJ, Lin WN (1970) Interfacial energy and structure in fcc metals and alloys. Philos Mag 22(177):515–542

Nabarro FRN (1967) Theory of crystal dislocations. Clarendon, Oxford, UK

Nabarro FRN (1969) Dislocations in a simple cubic lattice, In: Argon AS (ed) Physics of strength and plasticity. The MIT Press, Cambridge, MA, p 97
Romonov AE, Kolesnikova AC (2009) Application of dislocation concept to solid structures. Prog Mater Sci 54:740–769
Volterra V (1907) Sur l'equilibre des corps elastiques multiplement connexes. Ann Ecole Norm Super 24:400
Weertman J, Weertman JR (1992) Elementary dislocation theory. Oxford University Press, Oxford, UK

Planar Defects: Crystal Interfaces

Contents

Abstract

Aspects already discussed of stacking faults on {111} planes in fcc crystals and the emission of dislocations from grain boundaries in fcc polycrystals discussed in chapter "Line Defects: Dislocations in Crystalline Materials" have shown the dissociation of dislocations forming varying extensions of stacking faults. In fcc materials, stacking faults are characterized by partial dislocations, and similar dislocation configurations are characteristic of twin boundaries on {111} planes. Grain boundaries can also be shown to be composed of dislocation arrays, and interfaces separating different crystal regimes (interphase boundaries) as well as the accommodation of misfit and misfit strains between different phases can also be viewed as misfit dislocation arrays. In this chapter, these planar arrays are discussed in terms of their fundamental structures or microstructures and the corresponding or associated interfacial free energies. In some cases, where applicable, disclination concepts are also applied.

Introduction

The concept of planar defects applies to any interface, even for small second-phase inclusions, since, in an ideal case, an incremental area can be considered small enough that the curvature is inconsequential. In some sense, curvature is an important issue because a plane surface and a curved surface ideally constitute

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_16

energetically different regimes. There are of course examples of primarily planar interfacial regimes. As will be noted below, stacking faults and twin boundaries in face-centered cubic (fcc) metals and alloys are prominent examples. Some grain boundaries exhibit prominent planar features in some circumstances as well. Many precipitate or second-phase platelets which are coincident with specific crystal planes associated with the host crystal (or matrix) are also often ideally planar.

Stacking Faults and Twin Boundaries

Stacking faults and twin boundaries are particularly prominent in fcc metals and alloys and to some extent in diamond cubic structures. Stacking faults are also prominent in some hcp structures. In bcc structures, stacking faults and twin boundaries are far less common, even though both stacking faults and twin boundaries in bcc occur on the same {112} planes. However, since the primary glide planes in bcc are {110} planes, there is not a simple coherence between sessile dislocations on {110} and the dissociation of sessile dislocations on {112} to form stacking faults or twins.

In fcc, the close packing of {111} planes can be extended by stacking successive layers so that atoms occupy the interstitial regions of the previous layer: ABCABC, etc. For hcp, this stacking involves only two close-packed layers: ABABAB, as implicit, for example, in the graphite structure shown in ▶ Fig. 4 of chapter "Chemical Forces: Molecules." Stacking faults in the fcc sequence can occur simply either by inserting an A, B, or C layer or by extracting a layer. The former produces an *extrinsic* stacking fault, while the latter produces an *intrinsic* stacking fault. Since A, B, and C denote {111} planes, the twin can be represented by simple mirror symmetry in this notation: ABCABCBACBA, etc. Figure 1a illustrates these "defects" in this simple stacking sequence, while Fig. 1b represents them as an interface viewed edge on for an fcc crystal in the (110) orientation. The notation in Fig. 1b shows the $(\bar{1}11)$ plane as the fcc stacking-fault or twin plane, which corresponds to the $[1\bar{1}2]$ trace direction. In the representations in both Fig. 1a and b, the twin or the faults preserve close packing and there is no distortion of nearest-neighbor bonds. When viewed in projection in the TEM, there is a notable distinction between an intrinsic stacking fault and a coherent twin boundary as illustrated in Fig. 2. Since the coherent twin boundary at I (T) in Fig. 2 is a true interface separating the matrix (100) from the twin, the contrast fringes show systematic absorption, and fringes are added at the surface. In contrast, the intrinsic stacking fault at I (SF) in Fig. 2 is a simple planar defect on (111), and the fringe contrast is distinct from the twin boundary. Thickness changes are accommodated within the fault-plane contrast.

It is difficult to measure the stacking-fault free energy from extended partials or very long stacking faults (as in Fig. 2) because they are not equilibrium configurations required for ▶ Eq. 20 in chapter "Line Defects: Dislocations in Crystalline Materials". However, for extended nodes on {111} planes even in a thin film, the equilibrium radius can be reasonably determined and the stacking-fault energy

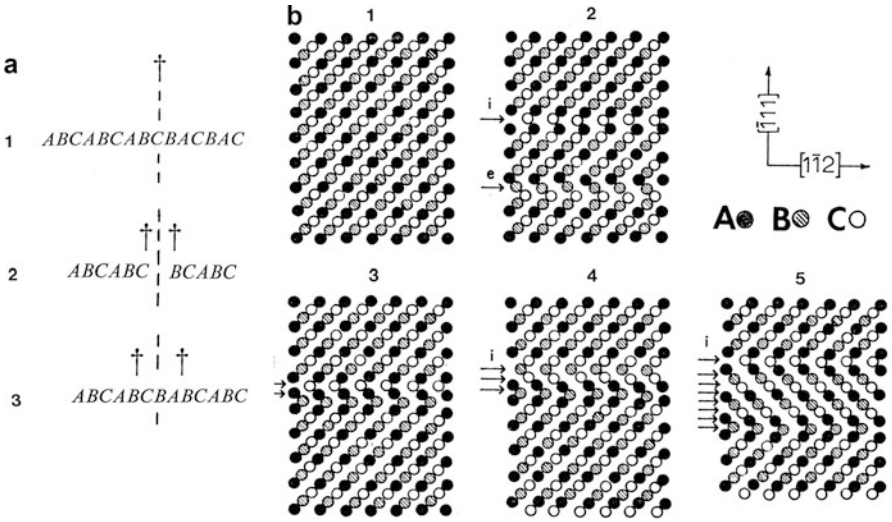


Fig. 1 (**a**) Stacking sequences for a coherent {111} twin (1), an intrinsic stacking fault (2), and an extrinsic stacking fault (3). The dagger denotes the interface plane, while the *dashed line* shows the fault center. In (2), the A plane is removed while in (3) the B plane is inserted. The extrinsic stacking fault in (3) is equivalent to two twin planes separated by two atomic layers. (**b**) View of {111} planes in (110) surface. (1) Perfect lattice. (2) Intrinsic (i) and extrinsic (e) faults. (3) and (4) Overlapping intrinsic faults forming narrow twins. (5) Wide twin ((**a**) is adapted from Hirth and Lothe (1968))

measured by ▶ Eq. 21 in chapter "Line Defects: Dislocations in Crystalline Materials". In addition, the variation of the stacking-fault free energy with temperature can be observed by the change of the extended node radii as the temperature is varied in a heating stage in the TEM as shown in the sequence of Fig. 3.

It can be observed in the context of simple packing geometry in Fig. 1 that the interfacial free energy can be approximated by

$$[\gamma_{SF} \cong 2\gamma_{tb}]_T \tag{1}$$

which implies that the intrinsic and extrinsic stacking-fault free energies are roughly the same, and approximately to the coherent twin boundary free energy, γ_{tb}. Equation 1 also indicated that the energies correspond to the same (equilibrium) temperature, T. Unfortunately, twin boundary free energies are not directly measured as in the case of stacking-fault free energies, and while, as illustrated in Fig. 3, relatively low temperatures can influence the stacking-fault free energy, coherent twin boundaries in fcc metals and alloys are usually equilibrated at temperatures near the melting point. Consequently, the energies must be adjusted by a knowledge of $d\gamma_{SF}/dT$ and $d\gamma_{tb}/dT$ (the corresponding temperature coefficients). From the first principles (thermodynamic principles) for pure metals (Murr 1975),

$$\frac{d\gamma_I}{dT} = -S_I, \tag{2}$$

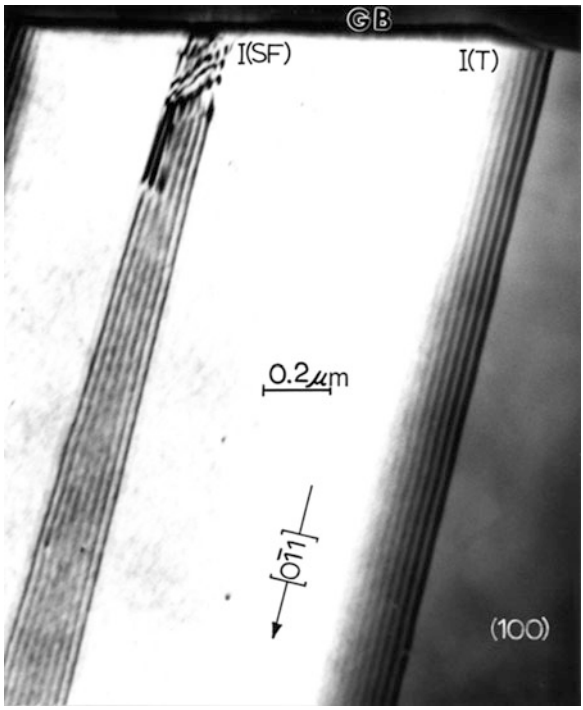


Fig. 2 Stacking fault (I (SF)) and coherent (111) twin boundary (I (C)) in copper intersecting a grain boundary at GB. The grain surface orientation is (100)

where γ_I is the interfacial free energy and S_I is the interfacial entropy. This often applies to many alloys where there is no interfacial segregation. However, in some alloys prone to elemental segregation, such as the Ag-Sn system in Fig. 3 (Ruff and Ives 1973), the interfacial free energy (in this case the stacking-fault free energy) can increase with increasing temperature.

Correspondingly, $d\gamma_I/dT = +$ (S). This can occur because Eq. 2 becomes

$$\frac{d\gamma_I}{dT} = -(S_I - \Gamma_i S_i), \tag{3}$$

where Γ_i is the interfacial excess of the ith component and S_i is the partial molar entropy of the ith component. In the case of the Ag-Sn system in Fig. 3, only one component will form an excess on the stacking fault, and as this excess product becomes large, the temperature coefficient, $d\gamma_{SF}/dT$, becomes positive as indicated by substituting the extended node radii into ▶ Eq. 21 in chapter "Line Defects: Dislocations in Crystalline Materials".

The energy of other internal interfaces, such as coherent and noncoherent twin boundaries and phase boundaries, can only be measured when corresponding grain boundary and surface free energies are known. In many instances, only interfacial free energy ratios can be measured by examining the intersections of interfaces equilibrated at specific temperatures: $(d\gamma_{I(1)}/d\gamma_{I(2)})_T$. For example, coherent twin boundary free energy/grain boundary free energy ratios can be determined by

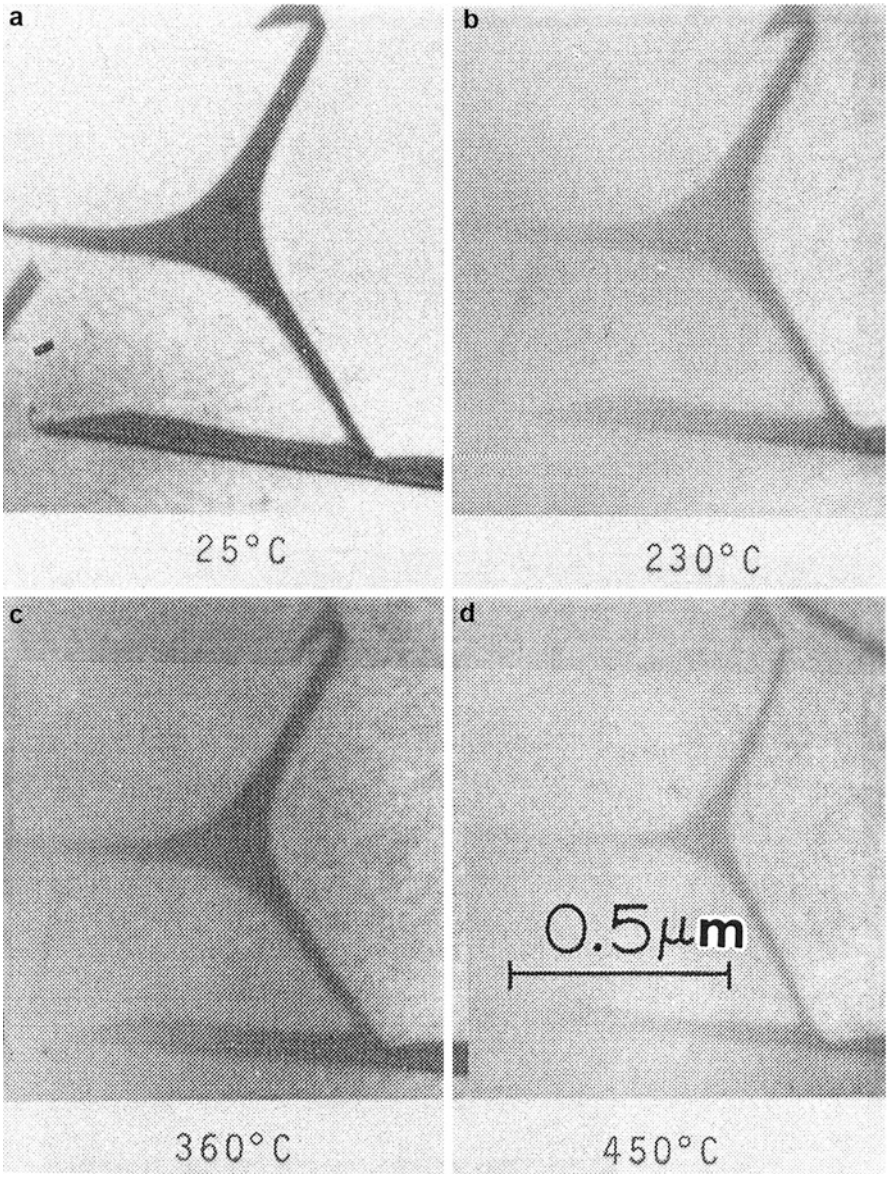


Fig. 3 Variation of dislocation (stacking fault) extended node radius (R) with temperature using a heating stage in the TEM for Ag-9 at.% Sn. The (111) plane is in the plane of the image (Courtesy of Dr. A. W. Ruff)

considering twin/grain boundary intersections as illustrated in Fig. 4. Here the grain boundary free energy is resolved as a vector parallel to the boundary plane or within the plane forming an equilibrium energy configuration where the resolved coherent twin boundary free energy vectors intersect. In the twin/grain boundary system

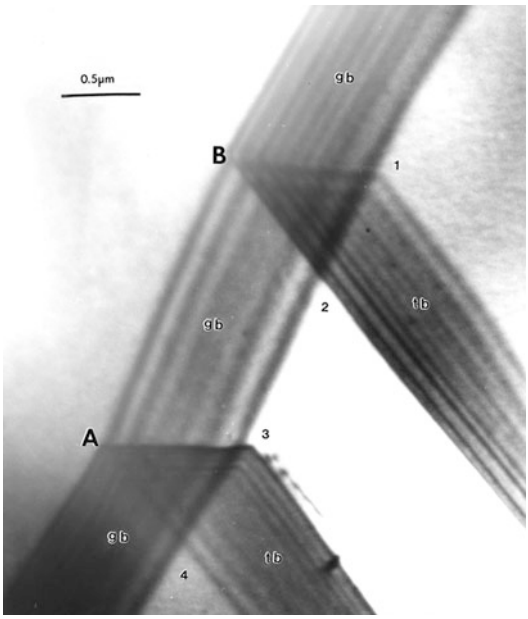


Fig. 4 Coherent twin boundary/grain boundary intersection in fcc 304 stainless steel. The matrix (grain) orientations on both sides of the grain boundary are (110). The twin plane is (111) (tb). Intersections A and B are at the top of the thin film. Twin boundary/grain boundary angles are denoted 1, 2, 3, 4, which are related to the true dihedral angles, Ω_1, Ω_2, Ω_3, Ω_4

shown in Fig. 4, there is a resolved torque which pushes the junction outward at B and pulls the grain boundary in at A, where A and B represent the twin plane/grain boundary plane junctions at the top of the thin foil section. By measuring the true dihedral angles, Ω_i, at 1, 2, 3, and 4 in Fig. 4 as described in detail by Murr (1975), the twin boundary free energy/grain boundary free energy ratio can be calculated from

$$\left[\frac{\gamma_{tb}}{\gamma_{gb}}\right]_T = \frac{1}{2}\sum_{i=1}^{4}(-\cos\Omega_i). \tag{4}$$

The associated torque terms arise because the two grains composing the intersection compete at the junctions,

$$\left.\begin{aligned}\sum M_A &= \frac{\partial\gamma_{gb}}{\partial\Omega}\sin\Omega_4 + \frac{\partial\gamma_{gb}}{\partial\Omega}\sin\Omega_3\\ \sum M_B &= \frac{\partial\gamma_{gb}}{\partial\Omega}\sin\Omega_1 + \frac{\partial\gamma_{gb}}{\partial\Omega}\sin\Omega_2\end{aligned}\right\}, \tag{5}$$

where ΣM represents the torque term or twisting of one orientation into the other at junctions A and B, respectively. These can also be calculated from the true dihedral angles, Ω_i, as the ratio $\Sigma M/\gamma_{gb}$ (Murr 1975).

The grain boundary free energy can be measured directly, along with the corresponding surface (solid-vapor) free energy using the so-called zero-creep method involving very small-diameter wires of the material in question. In this method, the dominance of solid surface tension or specific surface free energy is

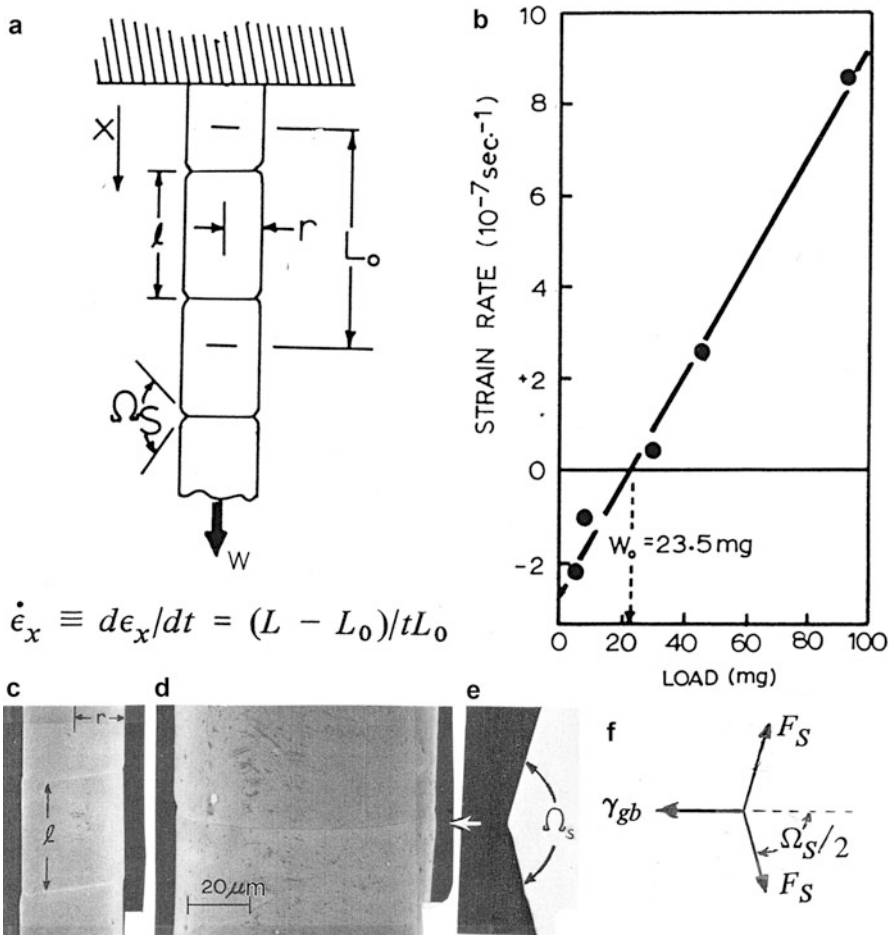


Fig. 5 Zero-creep method for measuring surface (F_s) and grain boundary (γ_{gb}) free energies. (**a**) Fine wires loaded with a series of weights (w) and crept for long times (t) in inert atmosphere at high temperature. The creep or strain rate is given by $\dot{\epsilon}_x$. (**b**) Experimental data for 80/20 Ni/Cr alloy at 1060 °C for 150 h in He. (**c**) and (**d**) are SEM views of crept wire. (**e**) shows a high magnification TEM shadowgraph of dihedral angle. (**f**) Equilibrium conditions for (**e**). For (**b**), $\gamma_{gb} = 756$ mJ/m^2; $F_s = 2160$ mJ/m^2 (Adapted from Murr (1975))

promoted by straining very thin wires at elevated temperature in an inert gas environment such as purified helium. Because the wire surface-to-volume dominates, the surface tension or resolved specific surface free energy, F_s, will act against an applied force, contracting the wire if the force is less than the "surface tension" or elongating if the force excels the surface tension. If the rate of straining (either $\pm\dot{\varepsilon}$) is plotted against weights attached to these wires, the balance weight w_o, at which the strain rate is zero, can be observed. Figure 5 illustrates this concept schematically and with actual observations. As shown in Fig. 5, as the

wire creeps ($\dot{\varepsilon} = d\varepsilon/dt$) at some elevated temperature near the melting point, the grain structure in the wire will minimize the internal energy forming a so-called bamboo structure where the grain boundaries became perpendicular to the wire axis, and equilibrium grooves form (by diffusional phenomena) where the grain boundary meets the surface. The balance equation corresponding to Fig. 5 becomes

$$w_o g = \pi r\left[F_s - \gamma_{gb}\left(\frac{r}{l}\right)\right], \tag{6}$$

where w_o is the balance mass, g is the gravitational constant, r is the wire radius, and l is the average grain length measured boundary to boundary as illustrated in Fig. 5. From the geometry of the grain boundary/surface intersection, particularly measuring the dihedral angle, Ω_s, it is observed that (Fig. 5f)

$$\left[\gamma_{gb} = 2F_s \cos\left(\frac{\Omega_s}{2}\right)\right]_T. \tag{7}$$

Consequently, from Eqs. 6 and 7, the grain boundary and surface free energies can be calculated at a constant (reference) temperature (T). Knowing the grain boundary free energy and having determined the twin boundary/grain boundary free energy ratio as in Eq. 4, the twin boundary free energy can be calculated.

Murr (1975) has made extensive measurements and tabulations of a wide range of surface, grain boundary, phase boundary, coherent and noncoherent twin boundary, and intrinsic stacking-fault free energies, including temperature coefficients. Table 1 lists a select number of these measurements for some fcc metals and alloys, while Table 2 lists stacking-fault and coherent and noncoherent twin boundary free energies for a select number of fcc metals and alloys adjusted to room temperature using appropriate temperature coefficients ($d\gamma/dT$) (Murr 1975). It can be noted in Table 3 that the noncoherent twin boundary free energies are much larger than the coherent twin boundary free energies. Figure 6 illustrates this energetic difference for noncoherent twin boundary and a coherent twin boundary intersecting the same grain boundary. The noncoherent boundary pulls the grain boundary producing a corresponding dihedral angle, Ω_N, which is considerably different from the dihedral angle associated with the coherent twin boundary, Ω_C, in Fig. 6. Of course these angles are the apparent angles and are not the true dihedral angles which must be geometrically corrected (Murr 1975). It can be noted in Table 3 that the progression of increasing interfacial free energies is $\gamma_{tb} \rightarrow \gamma_{TB} \rightarrow \gamma_{gb}$, where γ_{tb} and γ_{TB} are the coherent and noncoherent twin boundary free energies and γ_{gb} is the grain boundary free energy, all at the same reference temperature shown. It is of interest to compare ▶ Fig. 7c in chapter "Point Defects" with Tables 1 and 3 since the values in ▶ Fig. 7c of chapter "Point Defects" have been adjusted using the corresponding temperature coefficients ($d\gamma/dT$) to 670 °C. It is also quite remarkable to observe in Table 2 that Eq. 1 is almost exactly satisfied for all of the measured and corrected metals and alloys (where $d\gamma_{tb}/dT$ has been measured) (Table 1). This in itself provides some assurance of the validity of the concepts and methodologies discussed in a fundamental way.

Table 1 Interfacial free energies for select fcc metals and alloys (Data from Murr 1975)

Metal or alloy	F_s[a] (mJ/m^2)	T (°C)	dF_s/dT (mJ/m^2)	γ_{gb}[b] (mJ/m^2)	T (°C)	$d\gamma_{gb}/dT$ (mJ/m^2)	γ_{tb}[c] (mJ/m^2)	T (°C)	$d\gamma_{tb}/dT$ (mJ/m^2)
Aluminum	980	450	−0.40	324	450	−0.12	75	450	−0.07
Copper	1,780	925	−0.50	625	925	−0.10	24	800	−0.20
Gold	1,400	1,000	−0.43	378	1,000	−0.10	15	1,000	−0.008
Nickel	2,280	1,060	−0.55	866	1,060	−0.20	43	1,060	−0.02
Ni-20% Cr	2,160	1,060	–	756	1,060	–	19	1,060	–
Platinum	2,200	1,300	−0.60	660	1,300	−0.18	161	25	−0.04
Silver	1,100	950	−0.47	375	950	−0.10	8	950	−0.003
Stainless steel (type 304)	2,190	1,060	−1.76	835	1,060	−0.49	19	1060	+0.007

[a] F_s – specific surface free energy
[b] γ_{gb} – grain boundary free energy
[c] γ_{tb} – coherent twin boundary free energy

Table 2 Comparison of room temperature (25 °C)-adjusted intrinsic stacking-fault and coherent twin boundary free energies (Data from Murr 1975)

Fcc metal or alloy	Intrinsic stacking free energy, γ_{SF} (mJ/m^2)	Coherent twin boundary free energy, γ_{tb} (mJ/m^2)	Temperature (°C)
Nickel	128	64	25
Copper	78	40	25
Gold	45	23	25
Silver	22	11	25
Stainless steel (type 304)	21	12	25

Table 3 Comparison of coherent and noncoherent twin boundary free energies with grain boundary free energy at a common temperature for some FCC metals and alloys (Data from Murr 1975)

Metal or alloy	γ_{tb}[a]	γ_{TB}[b]	γ_{gb}[c] (mJ/m^2)	Temperature (°C)
Copper	21	498	623	950
Cu-5 % Al	17	180	560	950
Silver	8	126	377	900
Stainless steel (type 304)	19	209	835	1,060

[a]γ_{tb} – coherent twin boundary free energy
[b]γ_{TB} – noncoherent twin boundary free energy
[c]γ_{gb} – grain boundary free energy

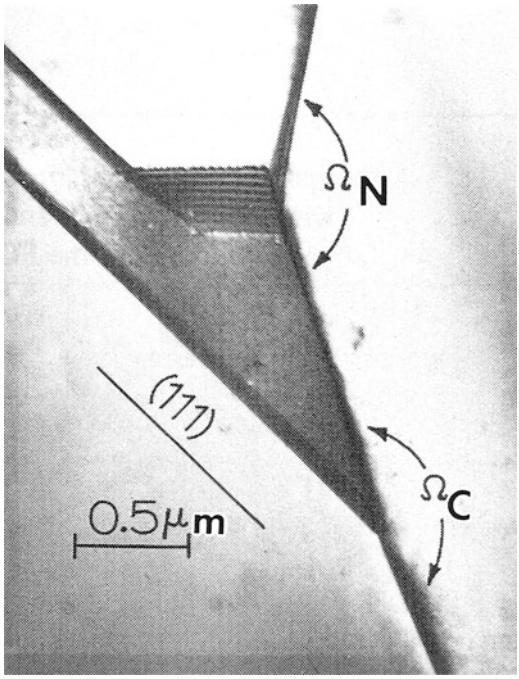


Fig. 6 Coherent and noncoherent twin boundary segment intersecting a grain boundary in 304 stainless steel at 1,060 °C. The apparent (dihedral) angles are denoted Ω_C and Ω_N, respectively (Adapted from Murr 1975)

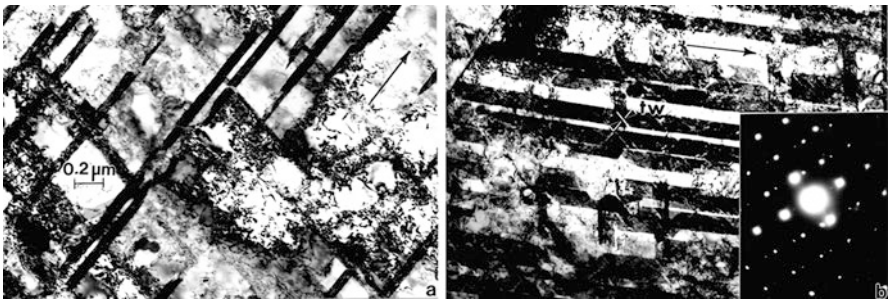


Fig. 7 Deformation microtwins in plane shock-loaded Ni. (**a**) 30 GPa shock pressure. Mixture of microtwins and dislocation cells. (**b**) 40 GPa shock pressure. *tw* denotes twin-fault intersection on different (111) planes. The surface orientation in (**a**) and (**b**) is (100). This is illustrated in the selected-area electron diffraction pattern insert in (**b**) which also exhibits twin spots. The magnification of (**a**) and (**b**) is the same and shown in (**a**). *Arrows* in (**a**) and (**b**) denote trace of (111) plane

Planar-Related Deformation Microstructures: Microtwins and Microbands

The origin of deformation-related microtwins or twin faults is rested in the schematic sequences of Fig. 1b which, for fcc metals and alloys, constitute overlapping arrays of stacking faults. In various modes of deformation, these overlapping arrays are characteristically irregular and form thin, coherent twins containing interval regions of irregularly stacked faults and even dislocations. In many deformation modes, there is a fairly well-defined slip – twinning transition, somewhat implicit in the progressive strain regimes shown for tensile uniaxial deformation in ▶ Fig. 18 of chapter "Line Defects: Dislocations in Crystalline Materials." More pronounced microstructure transition has been observed for dislocation/stacking fault and twin-fault transition in uniaxial, shock-wave compression of fcc metals and alloys.

▶ Figure 31 in chapter "Line Defects: Dislocations in Crystalline Materials" illustrates the progression of dislocation cell formation with increasing shock pressure in Ni. As will be seen a little later, the dislocation cell wall, while not a simple, planar interface, constitutes a low-angle interfacial region which, as illustrated in ▶ Fig. 31 of chapter "Line Defects: Dislocations in Crystalline Materials", consists of transition, somewhat implicit in the progressive strain regimes shown for tensile uniaxial deformation in ▶ Fig. 18 of chapter "Line Defects: Dislocations in Crystalline Materials." More pronounced microstructure transition has been observed for increasing dislocation density and decreasing dislocation wall thickness. There is also a small but incremental increase in the misorientation angle characterizing the cells on either side of the cell wall. In plane shock loading, there is a transition at a critical pressure of ~30 GPa to deformation twinning, which becomes more pronounced at increasing shock pressure. Figure 7 illustrates this phenomenon, in contrast to ▶ Fig. 31 chapter "Line Defects: Dislocations in Crystalline Materials". Figure 7b shows the intersection of twin faults to provide

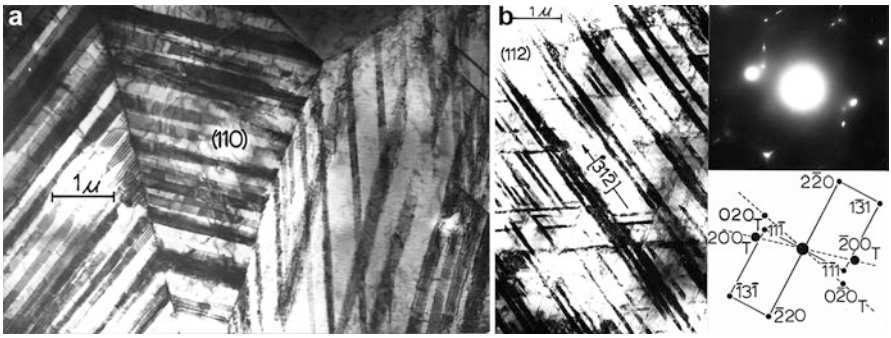
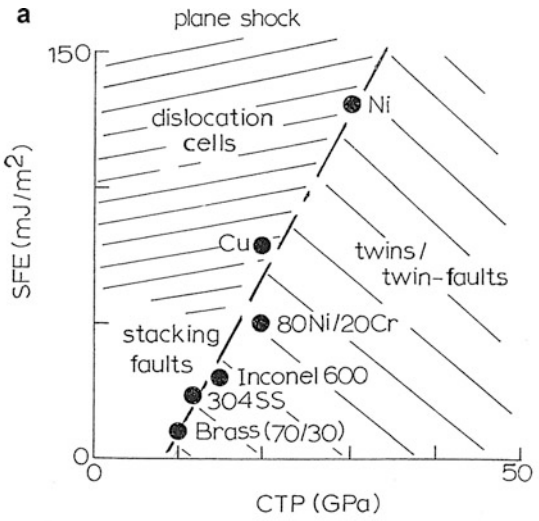


Fig. 8 Stacking faults and microtwins in plane shock-loaded type 304 stainless steel. (**a**) 12 GPa shock pressure. Stacking faults are intermixed with a few microtwins. (**b**) Microtwins following shock pressure of 30 GPa. The selected-area electron diffraction pattern insert and the corresponding diffraction net show twins spots for twins along $[31\bar{2}]$ in the $(1\bar{1}1)$ plane. The surface orientation is (112)

some measure of their thickness, which at (tw) is ~20 nm. The contrast along the twin fault in Fig. 7b is indicative of their irregular internal structure.

In contrast to ▶ Fig. 31 of chapter "Line Defects: Dislocations in Crystalline Materials" and Fig. 7 for Ni, which has a stacking-fault free energy of ~130 mJ/m^2 (Table 2), 304 stainless steel, with a much lower stacking-fault free energy of ~20 mJ/m^2 (Table 2), forms planar dislocation arrays and stacking faults at lower shock pressure up to about 12 GPa and then forms deformation twins or twin faults at higher shock pressures. This is illustrated in Fig. 8 for comparison with ▶ Fig. 31 of chapter "Line Defects: Dislocations in Crystalline Materials" and Fig. 7 for Ni. Figure 9a shows a plot of stacking-fault free energy and the critical twinning pressure for shock-loaded Ni and 304 stainless steel along with Cu and other lower stacking-fault energy alloys.

Figure 9b shows another deformation-related transition involving different interfacial microstructures defining microbands which also evolve from dislocation cell structure development in high stacking-fault free energy metals and alloys subjected to severe deformation in non-planar (uniaxial) modes. For example, impact cratering in copper and nickel and high stacking-fault free energy aluminum alloys such as 6061-Al forms microbands rather than twin faults or microtwins, which are characteristic of plane shock loading shown in Figs. 7, 8, and 9 (Murr 1988; Murr and Esquivel 2004; Quinones and Murr 1998; Murr 1981). Figure 10 shows a few examples of impact crater-induced microbands (Quinones and Murr 1998). It can be noted that in contrast to elongated dislocation cells, which often form in specific deformation or strain directions, microbands in fcc metals and alloys form in trace directions which are coincident with {111} planes, while the interfaces or dislocation boundaries composing the microbands are not perfectly coincident with {111}. Like dislocation cells or cell walls, microbands exhibit misorientations of only a few degrees and have a nominal width of ~500 nm (Fig. 10b). As illustrated in Fig. 9b and on comparing Figs. 7 and 10, microtwins and microbands in fcc are characteristic of different deformation modes or strain states.



b

METAL OR ALLOY	SFE (mJ/m^2)	PLANE SHOCK MICROSTRUCTURE		IMPACT CRATER (SPHERICAL SHOCK) MICROSTRUCTURE
Al 6061	150	microtwins*	→	microbands
Ni	130	microtwins	→	microbands
Cu	80	microtwins	→	microbands
304 Stainless	20	microtwins	↔	microtwins†
brass (70 Cu-30 Zn)	10	microtwins	↔	microtwins

*Some observations of intermixed microbands in heavily-deformed starting material.
†Some intermixed microbands (approximately 1 in 10 microtwins)

Fig. 9 (**a**) Stacking-fault free energy (SFE) versus critical twinning pressure (CTP) for fcc metals and alloys subjected to plane shock loading. (**b**) Matrix showing microtwin-microband transition in fcc metals and alloys deformed in plane shock and impact cratering

Figure 11 compares the principal features of microtwins (or twin faults) and microbands represented by Figs. 7, 8, and 10, in the same crystallographic context as shown schematically in Fig. 1b for an fcc (110) crystal or grain surface orientation. The trace direction for $(\bar{1}11)$ planes along $[1\bar{1}2]$ indicates the coherent twin or twin-fault planes and the orientation for the microbands. The microbrand interface is characterized by a double dislocation wall structure as described by Huang and Gray (1989) which can also be considered to be represented by a disclination dipole ($\pm\ \omega$).

Microbands have also been observed in the heavy rolling deformation of both fcc and bcc metals and alloys (Klemm et al. 2002; Luft 1991), although, in some cases, such as fcc brass (^{70}Cu-^{30}Zn), cold rolling can produce microtwins in multiple-pass uniaxial reverse rolling while microbands can form in high-strain cross-rolling. However, in bcc metals such as Mo, it is often difficult (without

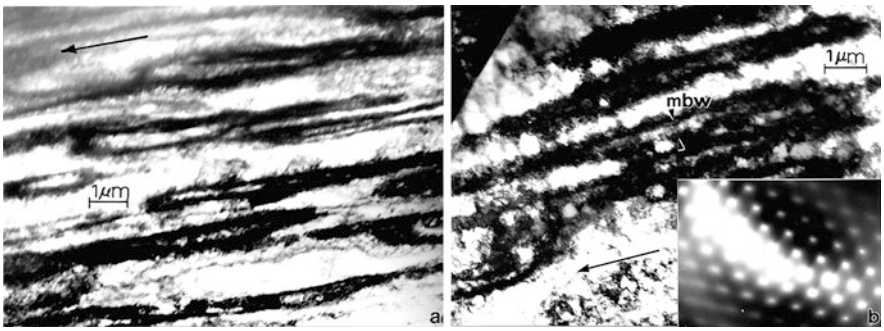


Fig. 10 Microbands formed in impact crater formation in (**a**) Ni impacted by a stainless steel projectile at 3.5 km/s (10.5 GPa shock). (**b**) Cu impacted by soda-lime glass projectile at 5.8 km/s. Selected-area electron diffraction pattern insert is (110). *Arrows* in (**a**) and (**b**) show trace directions for (111) planes

observing the diffraction patter) to distinguish microbands from microtwins since both are irregular, lenticular-like microstructures oriented along the trace of slip or twinning systems: {110} or {112}.

Figure 12 illustrates microtwin microstructures for shock-deformed Mo which are similar in appearance to the microbands shown in Fig. 10. Twins or microtwins in bcc consist of lamellar boundaries consisting of [111]/6 partial dislocations:

$$\left.\begin{aligned} \frac{a[11\bar{1}]}{2} &\rightarrow \frac{a[11\bar{1}]}{3} + \frac{a[11\bar{1}]}{6} \\ \frac{a[11\bar{1}]}{3} &\rightarrow \frac{a[11\bar{1}]}{6} + \frac{a[11\bar{1}]}{6} \end{aligned}\right\}. \tag{8}$$

These dislocations allow each (112) plane to shift over its neighbor in a direction parallel to $[11\bar{1}]$, but the shifting is not as regular as (111) planes in fcc. However, in some cases, twins or microtwins in bcc metals can become more regular, producing long segments coincident with (112) planes. This feature is illustrated for deformation microtwins in bcc Ta in Fig. 13, in contrast to Fig. 12. Figure 13a shows extensive (112) coincident microtwins, while Fig. 12b shows a mixture of coherent (112) twin segments and irregular, lamellar-type segments on other (112) planes similar to Fig. 12. Figure 13c shows the extensive twin diffraction corresponding to Fig. 13b.

Twins and microtwins can occur in many other materials and can accommodate misfit and growth phenomena as well as the provision for more efficient shear deformation as a deformation mechanism. Figure 14 shows microtwins in the basal plane (001) for orthorhombic, superconducting $YBa_2\ Cu_3\ O_7$ formed by shock deformation (Murr 1991). Similar twins can form to accommodate the variance in the lattice constant: a =3.83 Å, b = 3.89 Å. These twins, in contrast to those in deformed fcc metals and alloys, shown in Figs. 7 and 8, are somewhat irregular and forming lenticular structures similar to those shown for bcc Ta in Fig. 13a, although the twin planes in the superconductor are (110).

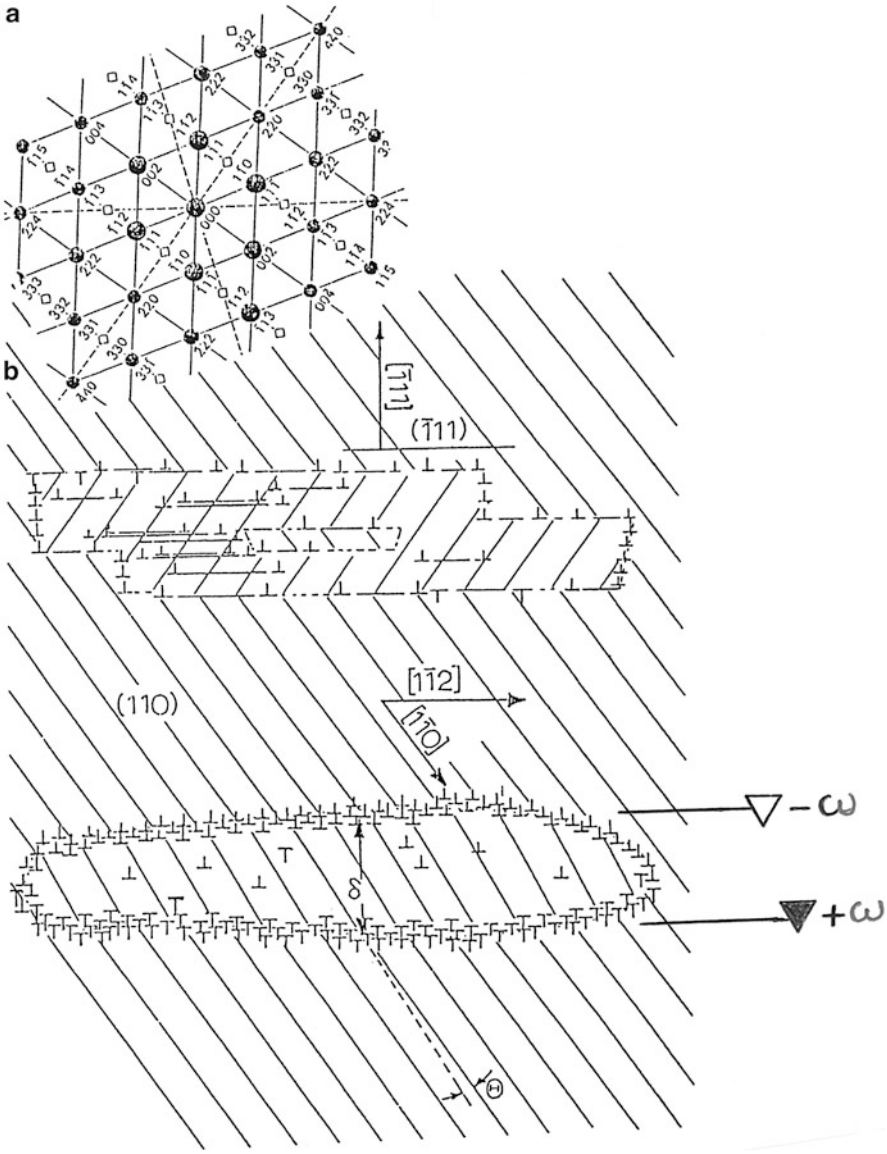


Fig. 11 Schematic representation for a deformation twin (**a**) and a microband (**b**) in the (110) grain surface orientation for an fcc crystal. The upper left insert shows a (110) diffraction net oriented appropriately. *Dotted lines* are (111) trace directions. $\pm$ ω represents a disclination dipole. Dislocations are indicated by $\perp$

In the context of interfacial structures or microstructures characteristic of planar defects, deformation structures or microstructures differ, as noted previously, by their irregularity, particularly irregularity in dislocation structures and arrangements. These are illustrated ideally in Fig. 11. We have already alluded to the similarities and

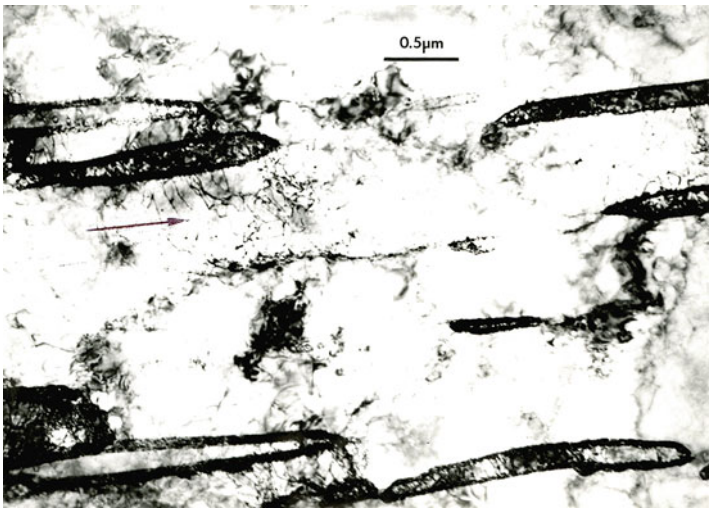


Fig. 12 Shock-wave-induced microtwins in bcc Mo. *Arrow* shows trace of (112) plane. 10 GPa shock pressure

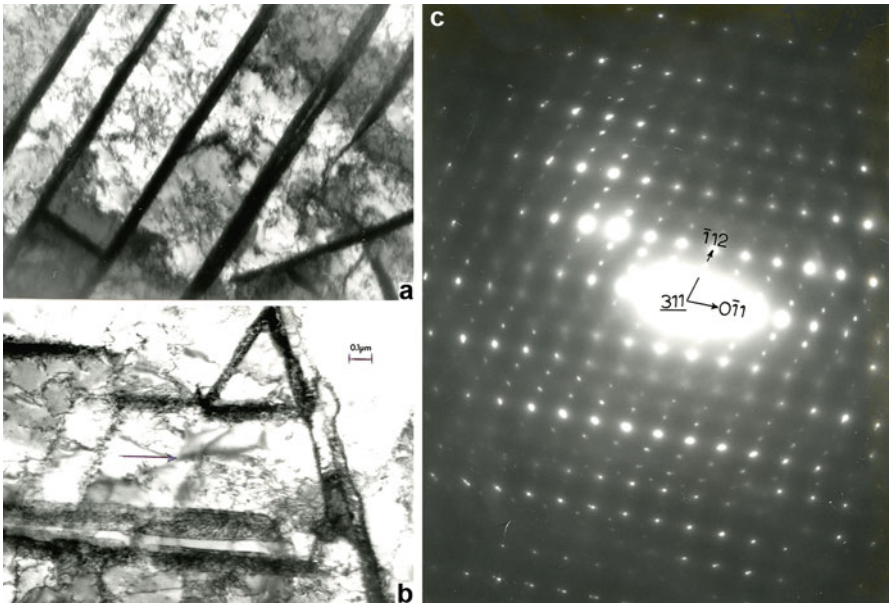


Fig. 13 Deformation microtwins in plane shock-loaded (45 GPa) bcc Ta. (**a**) (100) grain surface orientation showing twins coincident with four different {112} planes. (**b**) (311) grain surface orientation. *Arrow* shows $[0\bar{1}1]$ trace direction for (211) plane. Twins are also coincident with $(2\bar{1}1)$ and $(21\bar{1})$ planes. (**c**) Selected-area electron diffraction pattern corresponding to (**b**). Twin spots are at <112>/3 reciprocal lattice positions. Magnification is shown in (**b**)

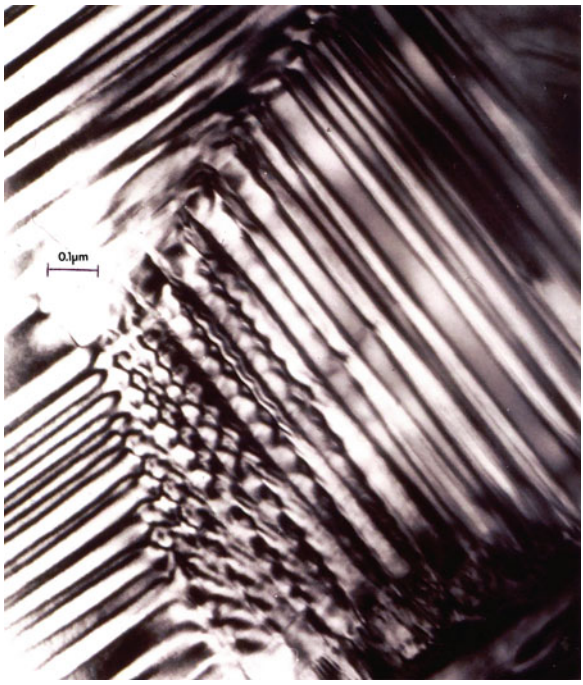


Fig. 14 Plane shock-wave-induced microtwins in $YBa_2Cu_3O_7$. 19 GPa shock pressure

differences of microbands and elongated dislocation cell structures where, aside from the coincidence of microbands, the dislocation wall structure (Fig. 11) will be more irregular than that for elongated dislocation cell walls. Similarly, these interfacial structures will differ from those characterizing microtwins in bcc metals as illustrated in Figs. 12 and 13.

We will now take a closer look at the interfacial structures and microstructures which characterize grain and other phase boundaries, which in some cases can be viewed as regular dislocation or dislocation arrays. These can be complicated by other more complex dislocation and ledge structures which themselves will be altered by deformation.

Grain Boundaries, Phase Boundaries, and Special Boundaries

The concept of a grain boundary or an interface separating two crystal portions can be conceptualized by a simple rotation of one crystal portion relative to another as illustrated in Fig. 15. Here, the interface is synonymous with a systematic Moiré rotation pattern which changes complexity with the angle of rotation, Θ_r. This feature might be considered to illustrate an entropic state representation characterizing the interface, with $S_I \propto \Theta_r$; $0° < \Theta_r < 45°$: $[-S_I \propto \gamma_I]_T$ (Eq. 2). Similarly and as illustrated in Fig. 15f–h, two crystals in perfect (100) registry with different lattice parameters, a, produce a special misfit ($(a_2–a_1)/a_2 \times 100$ %) interface even

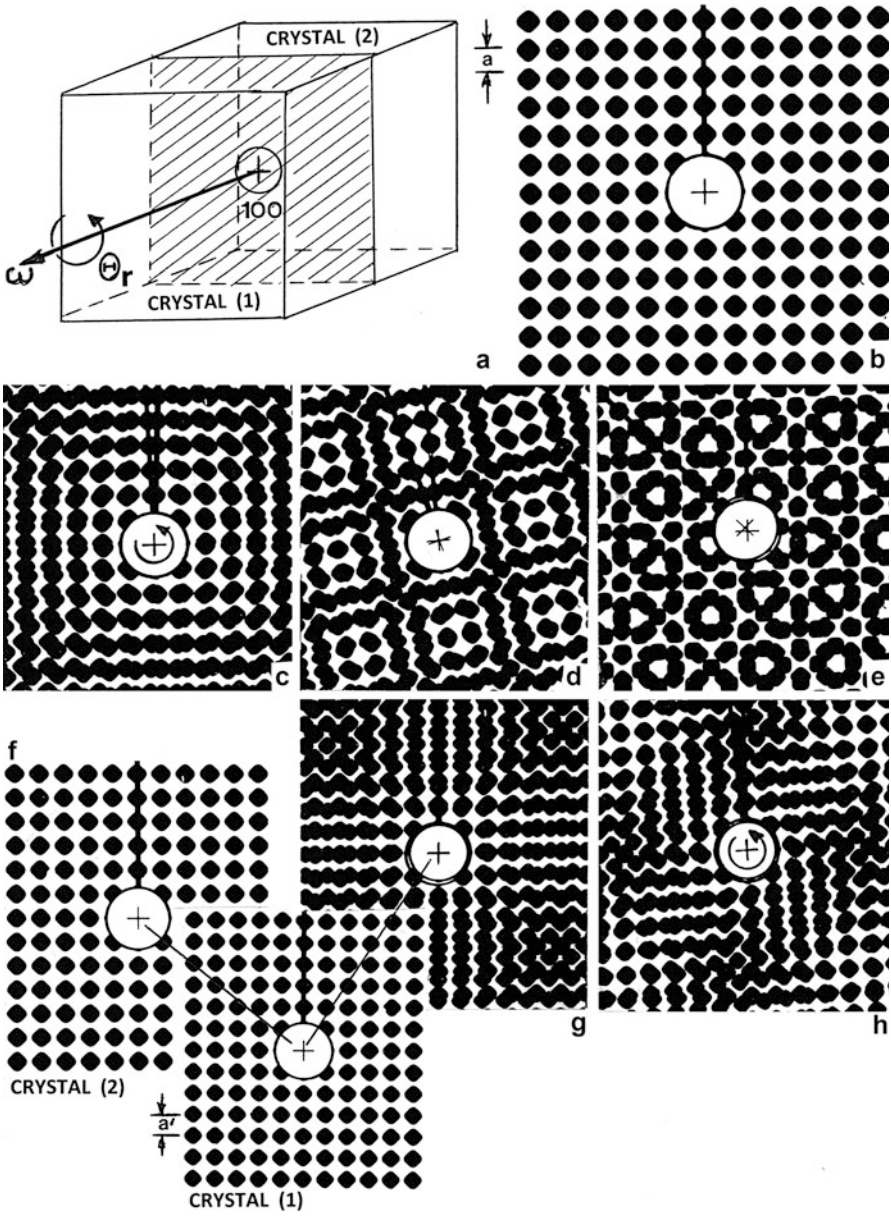


Fig. 15 Grain boundary structure generation by rotation of (100) atomic lattice arrays. (**a**) Single crystal containing virtual (*shaded*) grain boundary plane shown by atomic lattice in (**b**). (**c**) Lattice rotation by 5°. (**d**) Lattice rotation by 15°. (**e**) Lattice rotation by 45°. (**f**) and (**g**) show two different lattice/lattice parameters. (**h**) Rotation of lattices in (f) and (g) – crystal (1) relative to crystal (2) by 5°. Rotation axis is denoted **ω** in (**a**). Rotation angle is denoted Θ_r. Lattice parameters are indicated by a in (**b**) and a′ in (**f**)

without rotation and a correspondingly more complex interfacial "structure" with rotation, as illustrated in Fig. 15h. While this simple rotation-induced interface (Fig. 15c, d, e, and h) does not represent any systematic dislocation arrays, it might be represented by disclination arrays since the "defects" or deviations from perfect atom registry are rotation-induced and rotation specific. Similar so-called twist boundary structures can also be represented by screw dislocations.

Utilizing edge dislocation arrays, W. T. Read and W. Shockley, more than six decades ago (Read and Shockley 1950), introduced a similar crystal rotation or misorientation concept which is illustrated in Fig. 16. Here, the two crystal sections are oriented with regard to the interfacial plane (Fig. 16a) perpendicular to the reference system shown in Fig. 15a. Edge dislocations are introduced to systematically account for the misorientation (Θ) of one crystal relative to the other. This simple formalism assumes a symmetrical edge dislocation array to characterize the grain boundary (interfacial) structure, with Burgers vector $\mathbf{b}$, spaced a distance D (Fig. 16b) to accommodate a rotation of the two crystal lattices (or crystal grains A and B) about an orthogonal axis, $\boldsymbol{\omega}$. In the symmetrical case shown in Fig. 16b, c, the boundary trace ($\boldsymbol{\tau}$) and the normal to the boundary plane ($\mathbf{n}$) are orthogonal, and each bisects the total misorientation angle, Θ. From the grain boundary geometry illustrated in Fig. 16b,

$$\mathrm{D} = \frac{|\mathbf{b}|}{2\sin(\Theta/2)}, \tag{9}$$

or for small misorientation angles, Θ, $\mathrm{D} \cong \mathrm{b}/\Theta$. The corresponding Read-Shockley energy for this idealized grain boundary is given by

$$\gamma_{\mathrm{RS}} = \gamma_{\mathrm{o}}\Theta(\mathrm{A_o} - \ln\Theta) \tag{10}$$

where $\gamma_0 = \mathrm{Gb}/4\pi(1-\nu)$ for an edge dislocation (or tilt) boundary and, from Eq. 9, γ_{RS} is in units of energy per unit area ($\mathrm{mJ/m^2}$). This can be visualized in terms of the number of dislocations/unit length of boundary, which increases with decreasing D in Fig. 16: dislocation line energy/unit length of boundary energy/unit area. $\mathrm{A_o}$ in Eq. 10 represents the dislocation core energy. It is also shown in Fig. 16c that the symmetrical, edge dislocation (tilt) boundary can be represented by a wedge disclination array as shown in Fig. 16d.

It is often convenient to consider dislocation-structure boundaries to represent low-angle boundaries (small Θ or $\Theta < 15°$) and high-angle (of misorientation) boundaries: $\Theta > 15°$. If Eq. 10 is normalized to this condition, then we can let $\gamma_{\mathrm{RS}}/\gamma_{\mathrm{m}} = 1$ and $\Theta/\Theta_{\mathrm{m}} = 1$ and rewrite Eq. 9 as

$$\frac{\gamma_{\mathrm{RS}}}{\gamma_{\mathrm{m}}} = (\Theta/\Theta_{\mathrm{m}})\big(1 - \ln(\Theta/\Theta_{\mathrm{m}})\big), \tag{11}$$

which can be ideally represented by the sketch shown in Fig. 17. It can be observed in Fig. 17 (and by a simple interrogation of Eq. 9) that when D approaches the Burgers vector dimension ($\mathrm{D} \cong \mathrm{b}$), the dislocation cores will overlap. The core radius is not absolutely known, but is often approximated to be around 5b as noted

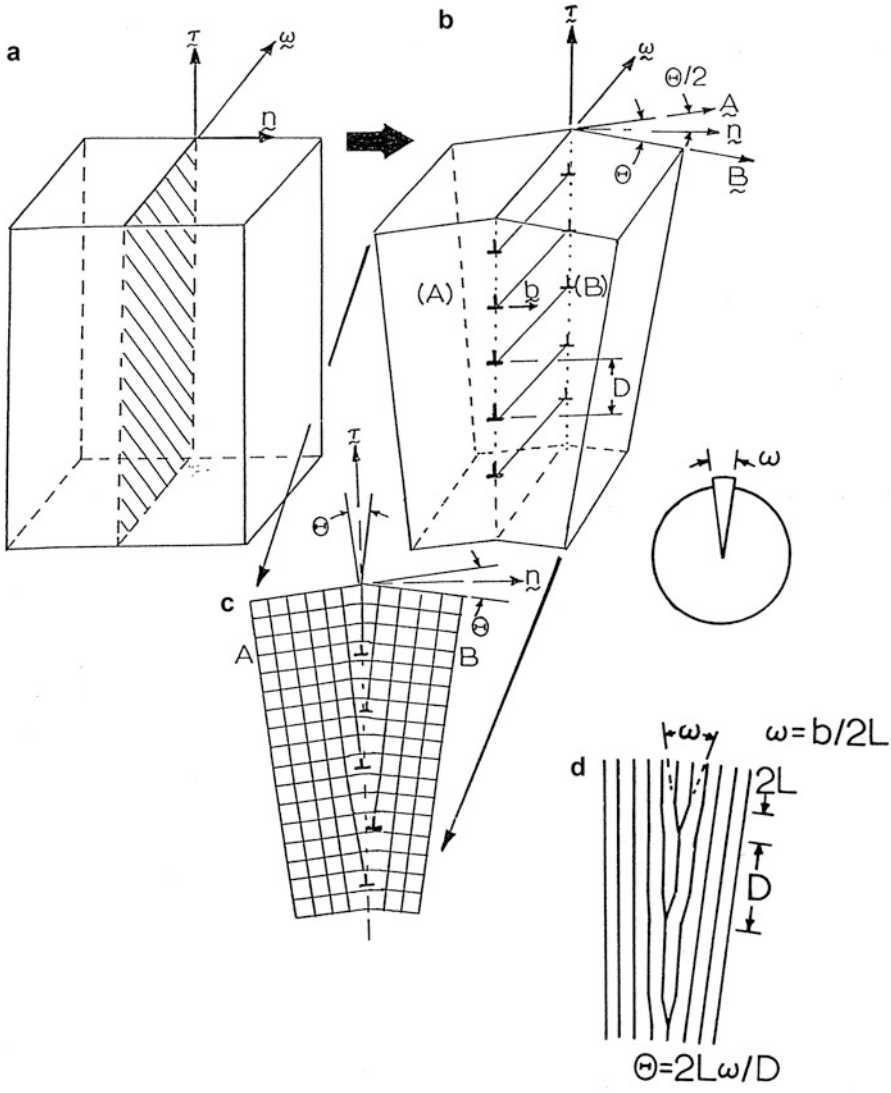


Fig. 16 Modified, symmetrical Read-Shockley (1950) dislocation boundary geometry and crystallography. (**a**) Single crystal with placement of virtual grain boundary plane (*shaded*) as in Fig. 15a. The boundary trace is denoted $\boldsymbol{\tau}$ and **n** is the normal to the boundary. **ω** is a rotation axis. $\boldsymbol{\tau}$, **ω**, and **n** are mutually orthogonal axes. (**b**) Symmetrical rotation of crystal A and B about **ω** to produce a dislocation boundary. **A** and **B** are crystallographic trace directions in crystal A and B, respectively. The total misorientation of **A** and **B** is denoted Θ. (**c**) A rendering of the lattice corresponding to a simple cubic (100) surface orientation of A and B as in Fig. 15b. Note the boundary plane (with trace $\boldsymbol{\tau}$) is perpendicular to the grain or crystal surfaces A and B, which in this rendering are identical: A = B. (**d**) A simple, symmetrical disclination boundary option for (**c**)

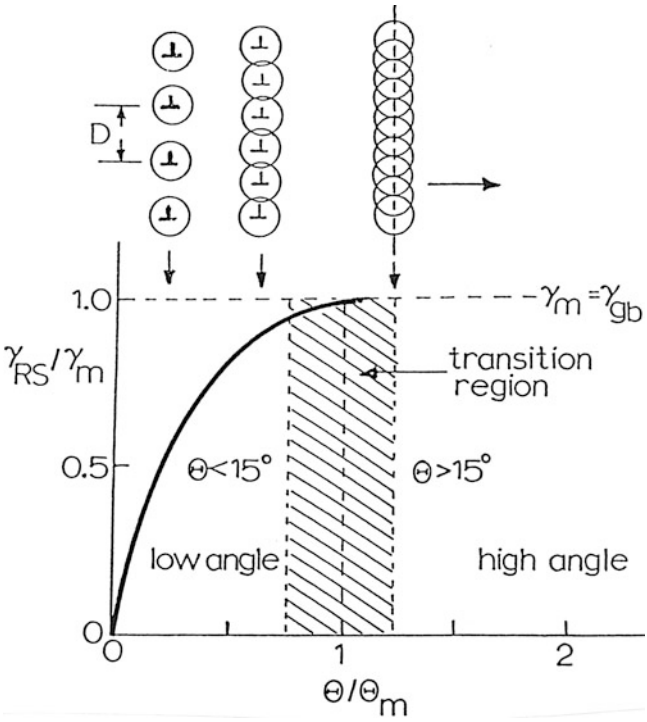


Fig. 17 Normalized plot of grain boundary free energy versus misorientation. Simple sketches of the corresponding dislocation boundary structure and dislocation core interaction/overlap evolution is shown above the plot. The usual, high-angle or high-energy grain boundary free energy is denoted γ_{gb}

in developing the line energies for dislocations in ▶ Eqs. 4 and ▶ 8 in chapter "Line Defects: Dislocations in Crystalline Materials". Assuming that the dislocation core overlap will become appreciable between about 3b to 6b, then the corresponding range of misorientation angles (Θ) will be roughly 10–19°. This corresponds roughly to the transition region in Fig. 17. Figure 17 shows that there is a strong correlation between grain boundary energy (specific interfacial free energy Murr 1975) and the misorientation angle, H, in the range $0° < \Theta < 15°$. These features were illustrated previously in ▶ Fig. 7 of chapter "Point Defects" in connection with interfacial free energies influencing corrosion sensitization in stainless steel. ▶ Figure 7 of chapter "Point Defects" also indicates cusps in the interfacial free energy (γ_I) versus Θ plot which correspond to so-called special or coincidence boundaries which, as consequence of coincidence, have reduced interfacial free energies (Murr 1975). Figure 18 illustrates this feature more specifically in terms of simple coincidence models (Fig. 18a, b) and calculated grain boundary free energy (γ_{gb}) versus misorientation angle (Θ) plots in Fig. 18c–e. Although the plotted values of γ_{gb} in Fig. 18c–e at 0 °K are much higher than experimentally measured and even adjusted interfacial free energy values ($d\gamma_I/dT$) for Table 1, they are in

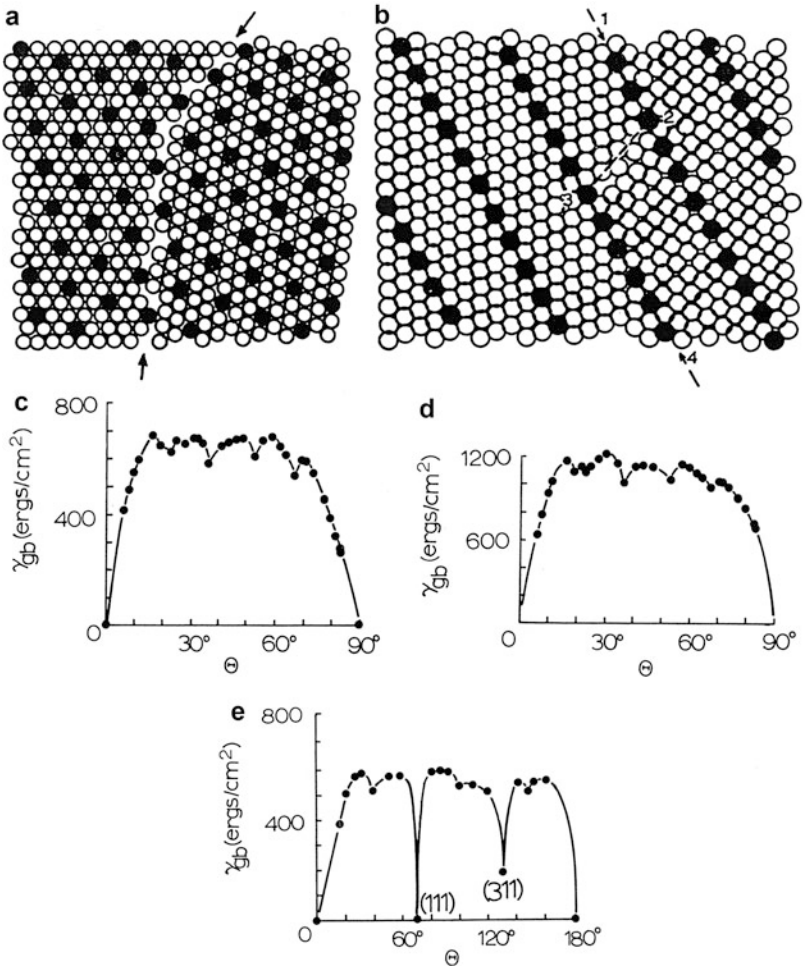


Fig. 18 Coincidence-site grain boundary models and calculated symmetrical tilt boundary energy versus misorientation plots. (**a**) Cubic (111) grain with misorientation of 38° corresponding to Σ7 boundary. (**b**) Cubic (110) grains with 50.5° misorientation corresponding to Σ11. (**c**) Aluminum <100> tilt boundaries. (**d**) Aluminum <110> tilt boundaries. (**e**) Copper <100> tilt boundaries. γ_{gb} is referenced to absolute zero (0°K) (After Murr (1975))

essentially the same ratio: $\gamma_{gb}(Cu) \cong 2\gamma_{gb}(Al)$ for the same orientation or <100> tilt boundaries. As illustrated in Fig. 18d, the calculated cusp for Σ3 coherent twin boundaries (111) corresponds to the experimental conditions implicit in ▶ Fig. 7 of chapter of "Point Defects." The higher-energy coincidence boundary for (311) shown in Fig. 18d can be composed with the noncoherent twin boundary free energy represented in ▶ Fig. 7 of chapter of "Point Defects" and implicit on comparing the coherent and noncoherent twin boundaries intersecting a grain boundary in Fig. 6.

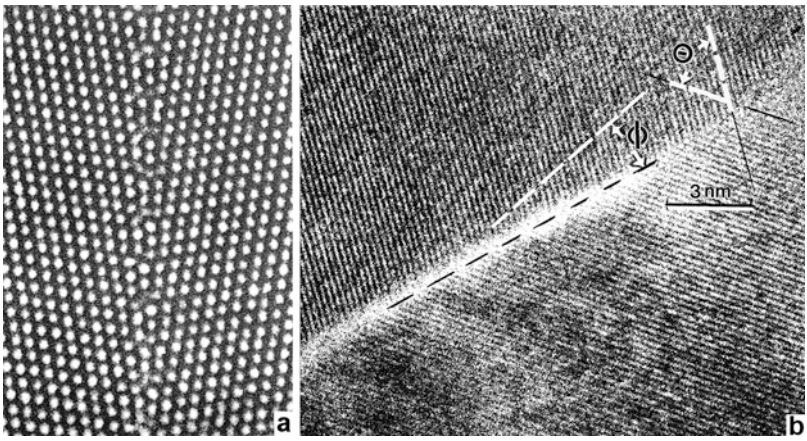


Fig. 19 Atomic (high-resolution) TEM image for symmetrical grain boundary in aluminum ($\Theta = 18°$) (**a**). TEM lattice image for asymmetrical grain boundary in aluminum ($\Theta = 56°$) (Courtesy of Ron Gronsky, UC Berkeley) (**b**). In (**b**), the grain boundary is indicated by the dark dotted line (GB). The symmetrical location of the boundary is indicated by the *dotted white line*. The asymmetry angle is denoted ϕ between the real grain boundary trace (GB) and the ideal trace (*white line*) $\phi = 14°$

Although Fig. 16 illustrated symmetrical grain boundaries characterized by regular or intrinsic dislocation arrays, many real grain boundaries, especially high-angle ($\Theta > 15°$) or correspondingly higher-energy grain boundaries, are neither symmetrical nor are they composed of only intrinsic grain boundary dislocations represented in Figs. 16 and 17.

Figure 19a compares an experimental (real) symmetrical grain boundary in aluminum resolved by atomic resolution imaging in the TEM, while Fig. 19b illustrates an asymmetrical grain boundary in aluminum resolved by lattice imaging in the TEM. As shown in Fig. 19b, the misorientation angle, Θ, is 56° in contrast to ~18° for Fig. 19a. The symmetrical grain boundary position which bisects this angle is denoted by the dashed white line, while the actual position of the grain boundary is shown by the dashed dark line in the actual boundary region. The angle between the ideal symmetry boundary (white) and the actual grain boundary (dark) is observed to be 14°. This angle is normally referred to as the asymmetry angle, ϕ, which in the context of the ideal symmetry array depicted in Fig. 16 is expressed by

$$D \cong \frac{b}{\Theta \cos \phi}, \tag{12}$$

which, in reference to Fig. 16, implies $0 < \phi < 45°$.

In real crystals, these two parameters, Θ and ϕ, are supplemented by four other parameters related to the inclination of the grain boundary relative to a reference surface θ (such as the grain surface in a thin film) as well as the orientation ((hkl) or [hkl]) of the grain (A and B) on either side of the grain boundary or a total of 6° of

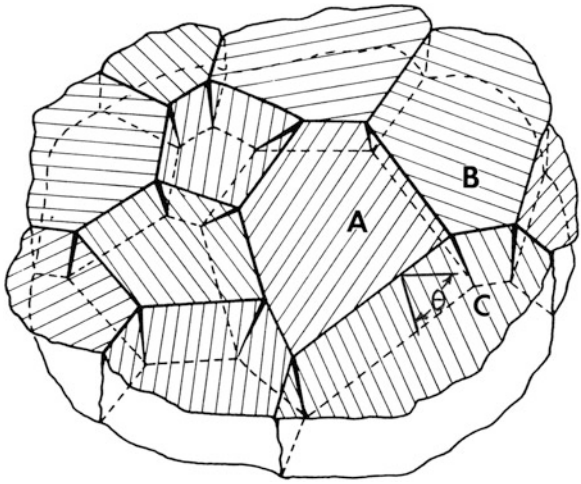


Fig. 20 Thin section representing polycrystalline grain structure. Grain boundary inclination is represented by θ. *A*, *B*, and *C* represent grain surface orientations: $(hkl)_A$, $(hkl)_B$, and $(hkl)_C$. The misorientation (Θ) can be represented by the angle between the "lattice lines" characterizing each grain

freedom characterizing the grain boundary crystallographically (or geometrically since the orientations (hkl) for grain A and B represent Cartesian coordinates, x, y, z, respectively, in A vs. B). Figure 20 illustrates such a section view of a polycrystalline material showing the intersection of grains A, B, and C, referred to as the triple junction of grain boundaries. For grains A and B having the same crystallographic surface orientation ($(hkl)_A = (hkl)_B$), the only remaining grain boundary parameters are Θ, ϕ, and θ, with Θ dominating the character of the grain boundary. However, as noted in the measurement of twin boundary/grain boundary intersection ratios (γ_{tb}/γ_{gb}) illustrated by Fig. 4, the true dihedral angles at triple junctions must be determined from the projected, apparent angles and the corresponding inclinations, θ, for the associated boundary planes (coherent twins and the grain boundary).

Figure 21 shows examples of intrinsic dislocation grain boundaries in projections of the grain boundary planes in bright-field TEM images, in contrast to the normal and high-resolution grain boundary structures shown in Fig. 19. In Fig. 21b, the dislocation spacing (d_1, d_2, and d_3 corresponding to D in Fig. 16) in the grain boundary segments averages ~40 nm in contrast to the intrinsic grain boundary dislocation (IGBD) spacing of ~10 nm in Fig. 21a. This corresponds, ideally, to misorientations of less than 1° in both Fig. 21a and b. This is essentially confirmed in the crystallographic features shown in Fig. 21b for stainless steel where $a = 0.36$ nm and the Burgers vector is ~0.25 nm in Eq. 9.

For a symmetrical low-angle Θ tilt boundary shown ideally in Fig. 19a, Fig. 10 can be expressed approximately by (Murr 1975)

$$\gamma'_{gb}(\text{tilt}) = \frac{Gb\Theta}{4\pi(1-\upsilon)}\left[\frac{4\pi(1-\upsilon)\gamma_c}{Gb^2} - \ln\frac{\in\Theta}{b}\right], \tag{13}$$

where γ_c is the dislocation core energy and $\in$ is the core cutoff radius for Θ, expressed in radians. Murr et al. (1970) have measured average misorientations, Θ, of 1.7° and 1.9° in 304 stainless steel and nickel, respectively, utilizing direct

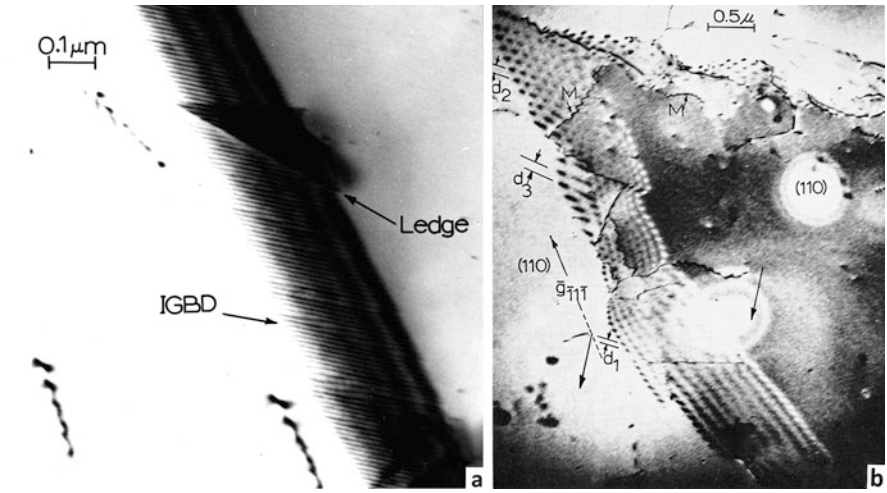


Fig. 21 Intrinsic grain boundary dislocations (IGBD) characterizing grain boundaries in stainless steel observed in the TEM. (**a**) Grain boundary projection in the TEM. (**b**) Two equivalent (110) grains separated by a dislocation boundary in projection. *Arrows* indicate $[1\bar{1}0]$ directions in each grain. Dislocation spacing is indicated by d_1, d_2, d_3. *M* denotes matrix (grain) dislocations

observations in the TEM similar to Fig. 21b. Ratios of γ'_{gb} (tilt/γ_{gb} were also measured from low-angle (tilt) boundaries intersecting high-angle grain boundaries similar to the intersections of a coherent twin or stacking-fault boundary intersecting a grain boundary as represented in Figs. 2 and 6, yielding values of 0.455 and 0.432 for nickel and stainless steel, respectively, at 1,060 °C. Correspondingly, from Table 1, values of γ'_{gb} (tilt) were determined to be 383 and 342 mJ/m^2 for Ni and 304 stainless steel, respectively. From Eq. 13, the average dislocation core energy can be written as

$$\overline{\gamma}_c \cong \frac{b}{\overline{\Theta}}\overline{\gamma'}_{gb}(\text{tilt}) + \frac{Gb\overline{\Theta}Ln}{4\pi(1-\upsilon)}\frac{\in\overline{\Theta}}{b}. \tag{14}$$

Substituting the average measured values of $\overline{\Theta}$ and $\overline{\gamma'}_{gb}$ (tilt) for nickel and 304 stainless steel and taking $\in \cong b$ into Eq. 14, the average dislocation core energy for Ni and 304 stainless steel was found to be $\sim 5.5 \times 10^{-5}$mJ/m and 5.9×10^{-5} mJ/m, respectively, representing roughly 6 % of the average dislocation self-energies (► Eq. 8 of chapter "Line Defects: Dislocations in Crystalline Materials"), in agreement with earlier theoretical predictions by Cottrell (1953) and Weertman and Weertman (1992).

It can be observed from earlier views of grain boundaries in ► Figs. 13, ► 14, and ► 18 of chapter "Line Defects: Dislocations in Crystalline Materials" that they generally have complex structures, including arrays of ledges and other grain boundary dislocations (GBDs). These dislocations are not an intrinsic part of the grain boundary and can often be thought of as allowing glide in the grain boundary

plane when grains deform, and such glide is often responsible for ledge production as well. The grain boundary itself is often a complex, transitional region which in fact has a thickness or width as opposed to a simple, mathematical dividing surface, often referred to as the Gibbs dividing surface. Figure 22 illustrates these concepts in simple schematics which, while representing these concepts, do not represent the complexity observed in real grain boundaries. Figure 23 illustrates the observation of intrinsic grain boundary dislocations (IGBDs) and GBD structure implicit in the schematic representation in Fig. 21b which is superimposed in Fig. 23. Figure 23a shows the regular (spaced) IGBDs and GBDs as well as dislocations in the grain or grain matrix outside of the grain boundary (circled area in Fig. 23). Figure 23b shows the same area tilted for reversed diffraction where the matrix or grain dislocations disappear, and the bottom of the thin foil represented by B is observed to be free of dislocation images, which in Fig. 23a are superimposed on the lower boundary plane image.

Although, in large grain size materials, the grain boundary width, Δt, in Fig. 23a, b is very small in relation to the grain size, D, as the grain size becomes much smaller, the grain boundary width can be only slightly smaller. In the extreme case of nanocrystalline materials, the grain size and grain boundary width become indistinguishable, and this poses special issues. Figure 24 illustrates these conditions, and we will address the special case of nanosize grain structures later in this presentation.

Figure 25 illustrates two examples of dislocation and ledge structures in situations where the grains composing the grain boundary have the same orientation, but the boundary itself is asymmetrical and with a correspondingly small misorientation, Θ. Figure 25a represents a misorientation, $\Theta \cong 6^\circ$, and shows regular, large ledges having a height of roughly 250 nm and similarly spaced apart. A very fine, nearly imperceptible arrangement of intrinsic grain boundary dislocations is observed along the arrow marked (I), and their spacing is approximately 100 nm. Additional, ledge-like features (white) do not interrupt these fine intrinsic structures but do interrupt the ledges previously described. In Fig. 25b, ledges are oriented close to the $(\bar{1}11)$ matrix (or grain) slip planes, while another set of ledges is oriented near the $(\bar{1}1\bar{1})$ planes. The two grains composing the boundary in Fig. 25b both have the common (112) orientation, and the asymmetry is similar to Fig. 25a, while the misorientation is the same: $\Theta \cong 6^\circ$. Figure 26 shows a grain boundary projection with three orientations of ledges corresponding almost exactly to the (111) slip planes in a (110)-oriented grain: (111), $(\bar{1}11)$, $(1\bar{1}1)$, and $(11\bar{1})$, with (111) and $(11\bar{1})$ sharing a common trace direction or corresponding ledge orientation. Finally, Fig. 27 shows a grain boundary with only very large ledges, one marked (LL) having an estimated ledge height from the boundary plane of ~200 nm. Since the effective Burgers vector of a grain boundary ledge can be represented by

$$\mathbf{B}_L = 2n\mathbf{b}\sin\left(\frac{\Theta}{2}\right), \tag{15}$$

where n represents the number of glide dislocations of Burgers vector, **b**, which can compose the ledge and be emitted as illustrated schematically in

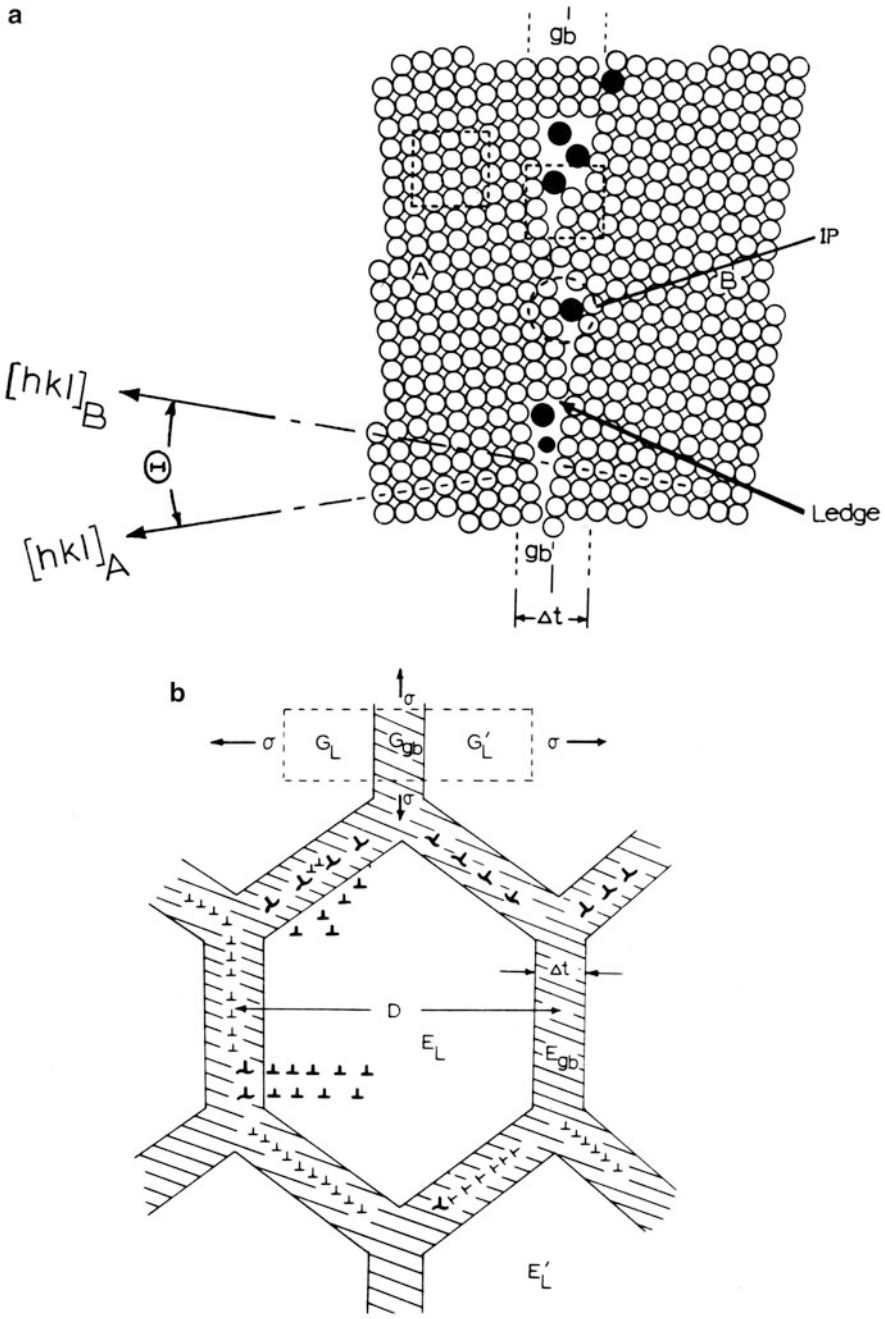


Fig. 22 Models for grain boundary structure. (**a**) Symmetrical grain boundary containing impurities and ledge structures, with a width of Δt. The grain boundary plane is indicated by gb. Common directions [hl] in grains A and B characterize the misorientation, Θ. (IP) denotes

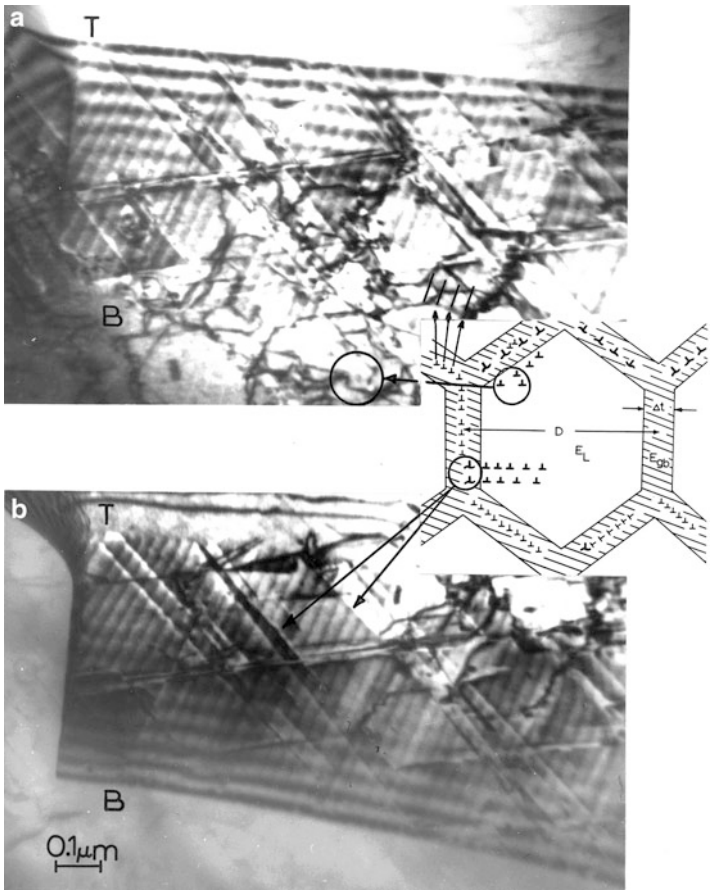


Fig. 23 Tilted views of grain boundary dislocation structures with schematic of Fig. 22b superimposed. (**a**) Intrinsic GBD arrays spaced ~30 nm are shown by *arrows*. The *circled area* shows matrix (grain) dislocations. (**b**) Tilted TEM view of (**a**) with different contrast operating reflection (**g**). GBD/ledges are shown by the *arrows*. T and B in (**a**) and (**b**) indicate foil (and grain boundary) *top* and *bottom*, respectively. Magnification is shown in (**b**)

► Fig. 14a–d of chapter "Line Defects: Dislocations in Crystalline Materials" to decompose the ledge itself. Figure 28 illustrates this single, coherent ledge emitting dislocations into the grain. Considering the ledge in Fig. 27, it could ideally produce ~1,000 dislocations emitted into the grain as shown schematically in Fig. 28.

Fig. 22 (continued) impurity. The dotted box area in grain A at upper left is a reference for the free volume element dotted in the grain boundary. (**b**) Grain boundary dislocations (⊥), intrinsic grain boundary dislocations (⊥) equally (regularly) spaced in the grain boundary region (Δt), and matrix dislocations (⊥) within the grain emitted from grain boundary dislocations (or ledges). The elastic (Young's) modulus in the grains are denoted E_L, E_L'; E_{gb} in the grain boundary regime. Corresponding shear stresses are denoted G_L; $G'_{L,}$ and G_{gb}

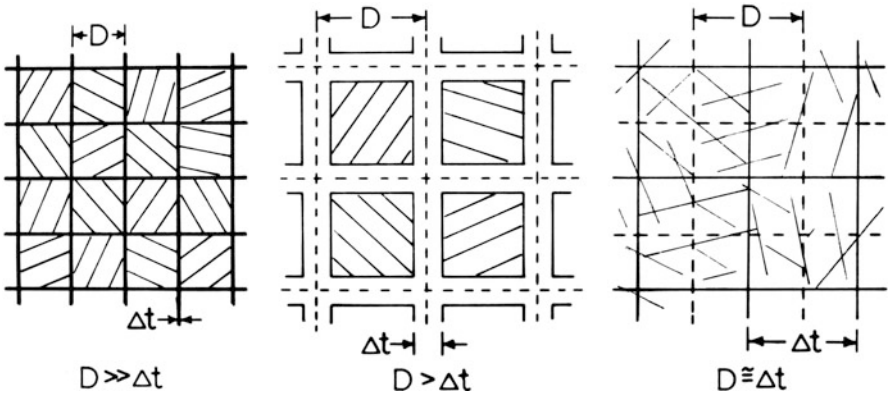


Fig. 24 Simple schematic views of grains with diameter, D, in polycrystalline and nanocrystalline (and glassy) materials ($D \cong \Delta t$). The interface phase (regime) width is denoted Δt. *Dotted lines* denote the Gibbs dividing surface. Grains are shown hatched

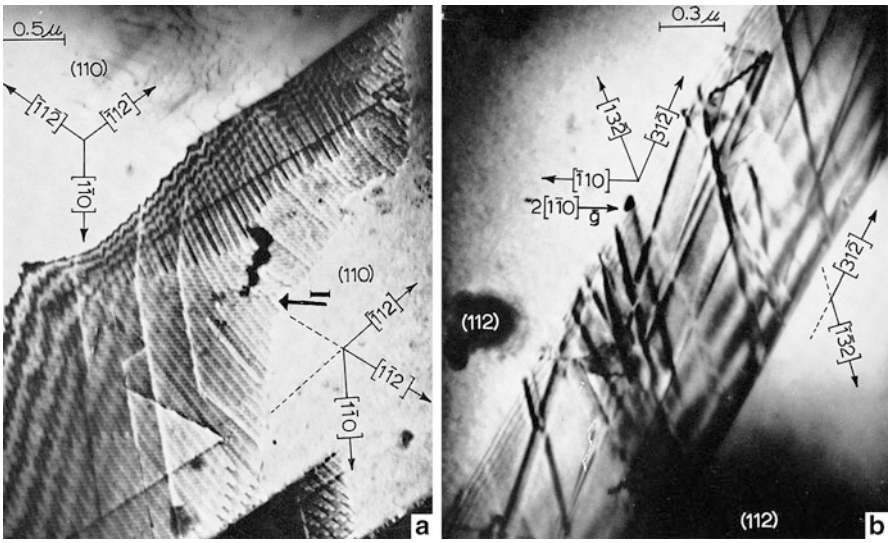


Fig. 25 Complex grain boundary and ledge structure in nickel. (**a**) Light cold reduction and anneal at 1,060 °C for 0.25 h. Misorientation of $[1\bar{1}0]$ directions is ~6°. (**b**) Dislocations and ledges. $[\bar{1}32]$ misorientation is ~6° as in (**a**) for $[1\bar{1}0]$. Note intrinsic grain boundary dislocations by sighting along arrow at (I) in (**a**)

In his original treatment of grain boundary ledges, J. C. M. Li (1961) noted that since the free energy for ledge formation is strongly dependent on grain boundary misorientation, Θ, ledges should, like intrinsic grain boundary dislocations, increase with misorientation, Θ. This was shown in fact in an Fe-0.15 Ti steel (Bernstein et al. 1972) and for nickel and type 304 stainless steel by Murr (1975) by

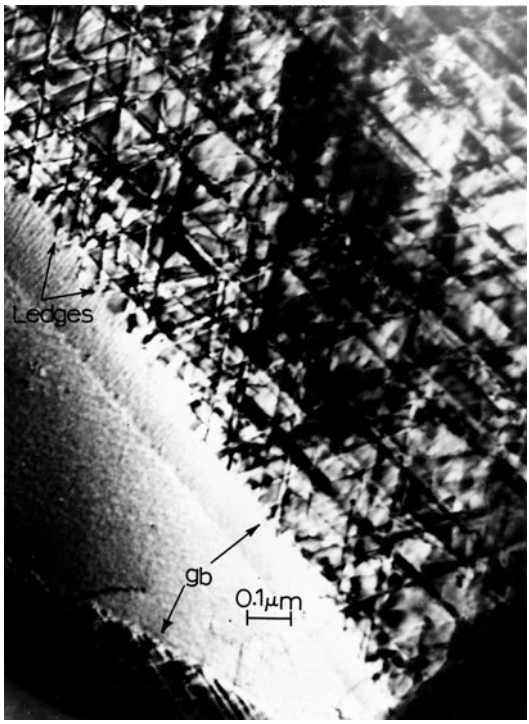


Fig. 26 Complex grain boundary (gb) ledge structure in a nickel sheet strained 0.2 %

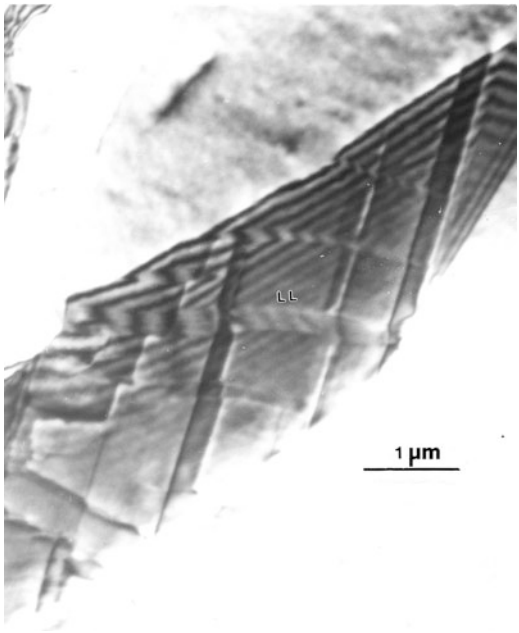


Fig. 27 Large ledge heights (LL) in a grain boundary in 304 stainless steel

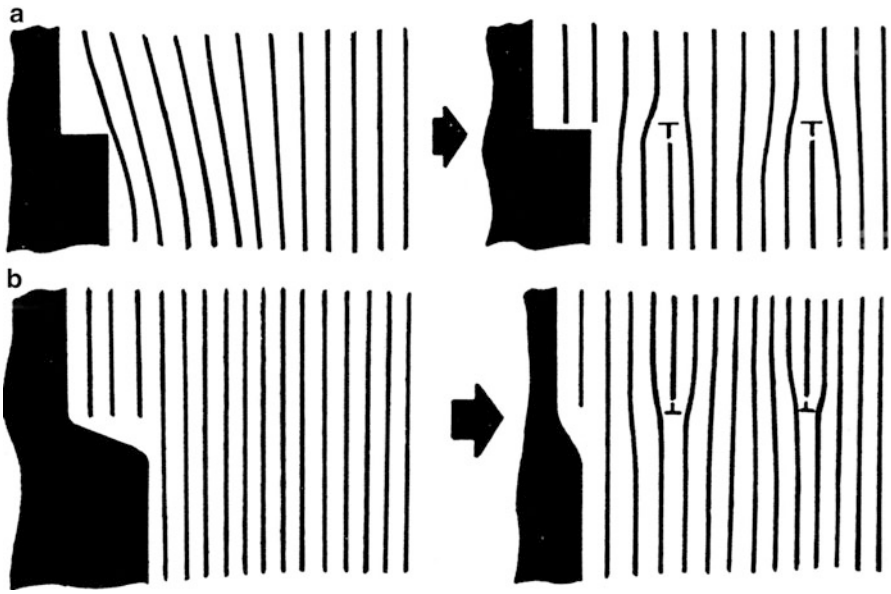


Fig. 28 Interfacial ledges as dislocation sources. (**a**) Single, coherent ledge emitting dislocations into the grain matrix. (**b**) Noncoherent ledge emitting dislocations (After Murr 1975)

measuring the number of ledges, m, per unit length of grain boundary versus misorientation, Θ. These results are reproduced in Fig. 29. Figure 30 illustrates a variety of grain boundary ledges in 304 stainless steel following thermomechanical treatment and light deformation. This illustrates that GBD structure, like the dislocation structure in the grains, can be thermomechanically manipulated. In this regard, grain boundary dislocations can participate in grain boundary ledge formation and the deformation of the grain boundary since they can glide in the grain boundary plane. Grain boundary structure will also change and become more complex with deformation. This is observed in the deformation sequence shown previously in ▶ Fig. 8 of chapter "Line Defects: Dislocations in Crystalline Materials," while Fig. 31 illustrates the grain boundary structure changes with strain in somewhat more detail.

In viewing the grain boundaries in Fig. 31, it becomes difficult to determine the degree of microstructure which is actually associated with the grain boundary and the dislocation structure occurring in the nearby grain (matrix) as a consequence of dislocation emission from grain boundary ledges. Figure 32 provides some indication of this phenomenon by viewing a coherent twin boundary in stainless steel which is perpendicular to the (110) grain surface. Viewed edge on, the emission and accumulation of dislocations and dislocation structure near the boundary plane are apparent and somewhat consistent with the observations of regular, high-angle grain boundaries in Fig. 5. The coherent twin boundaries in Fig. 32 contain ledges apparent, for example, in the schematic for microtwins or twin faults shown in Fig. 11.

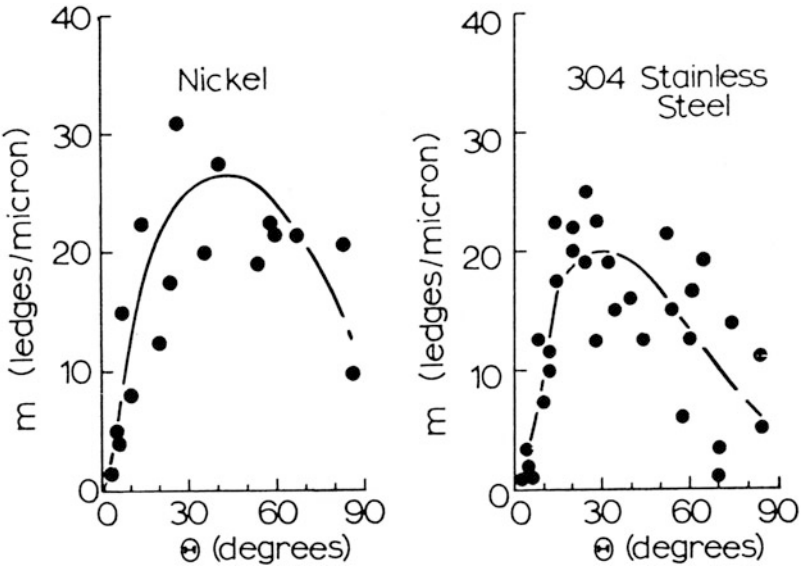


Fig. 29 Variation of grain boundary ledge density with misorientation in (**a**) nickel and (**b**) 304 stainless steel determined for identical (110) grain orientations composing the boundaries as observed by TEM. Misorientation was determined as the angle between common $[1\bar{1}0]$ directions

Marcinkowski (1972) has summarized the properties of grain boundary dislocations (GBDs) which are implicit in the many views provided in ▶ Fig. 8 of chapter "Line Defects: Dislocations in Crystalline Materials" and Figs. 21, 23, 25, 26, 27, 30, 31, and 32 as follows:

- *A GBD is distinct from a crystal lattice dislocation.*
- *A GBD can occur only within a grain boundary.*
- *The Burgers vector of a GBD may or may not lie within the grain boundary plane.*
- *The Burgers vector of the GBD is intimately related to the Burgers vector of the crystal lattice dislocations within each grain by the relation* $\boldsymbol{b}_I = \boldsymbol{b}_A + \boldsymbol{b}_B$, *where* $\boldsymbol{b}_I$ *is the interfacial or GBD Burgers vector and* $\boldsymbol{b}_A$ *and* $\boldsymbol{b}_B$ *refer to a dislocation in grain (A) heterogeneously shearing the boundary and emitted into grain (B)* (▶ Fig. 18c, d of chapter "Line Defects: Dislocations in Crystalline Materials").
- *The GBD type depends upon the nature of the grain boundary along with the nature of the dislocations cutting through the boundary.*
- *A GBD can be either sessile or glissile.*
- *GBDs can react with one another.*
- *GBDs can act as sources for crystal lattice dislocations* (by forming ledges).
- *The Burgers vector of a GBD is the smallest unit so that splitting of the GBD cannot occur.*
- *A GBD can also interact with impurities and precipitates* (Fig. 22a).

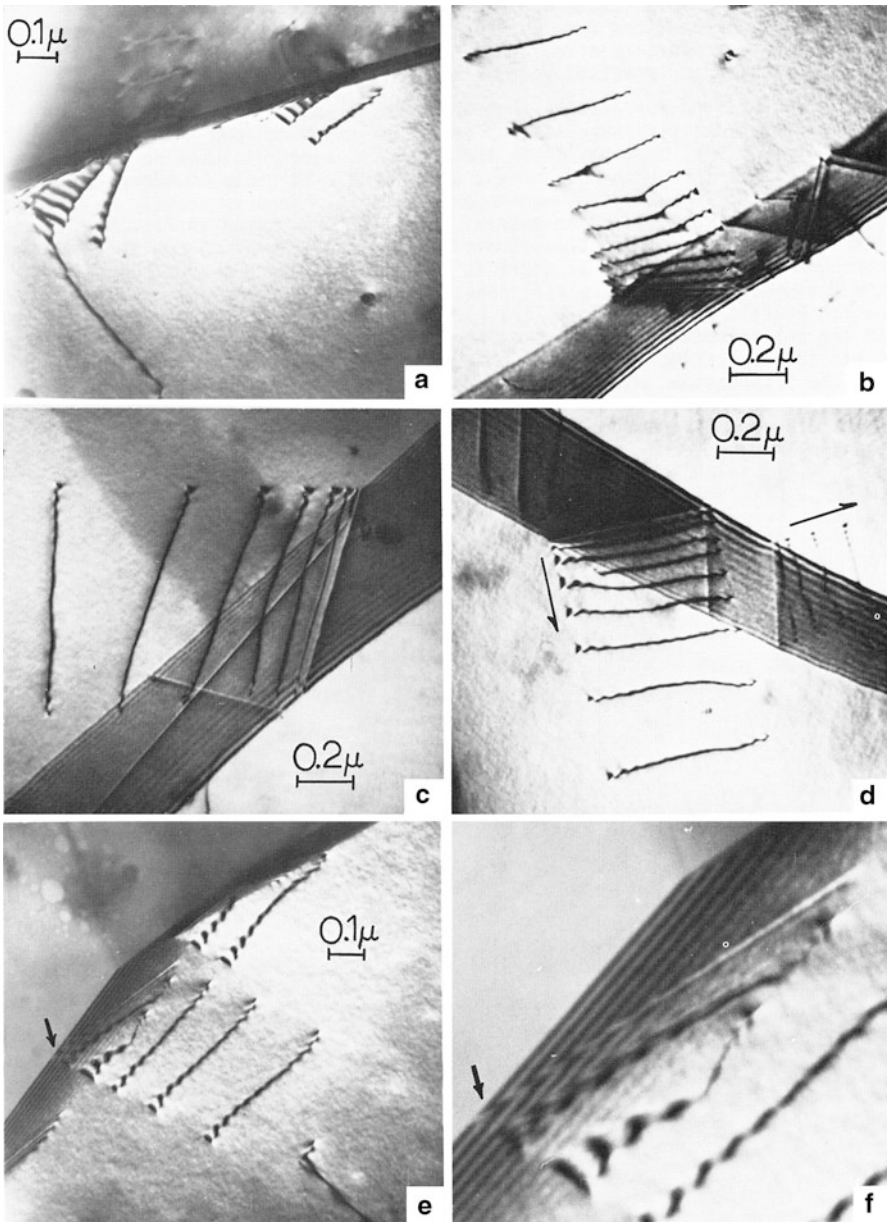


Fig. 30 Examples of grain boundary ledge dislocation sources following thermomechanical treatment and light deformation in 304 stainless steel (cold rolling by 30 %, annealed 1 h at 1,100 °C and water quenched); (**a**)–(**d**) show short dislocation profiles emitted from ledges. (**e**) Dislocation profile associated with large grain boundary ledge. (**f**) Magnified view of boundary ledge of (**e**) showing boundary fringe distortion indicative of large step (*arrow*)

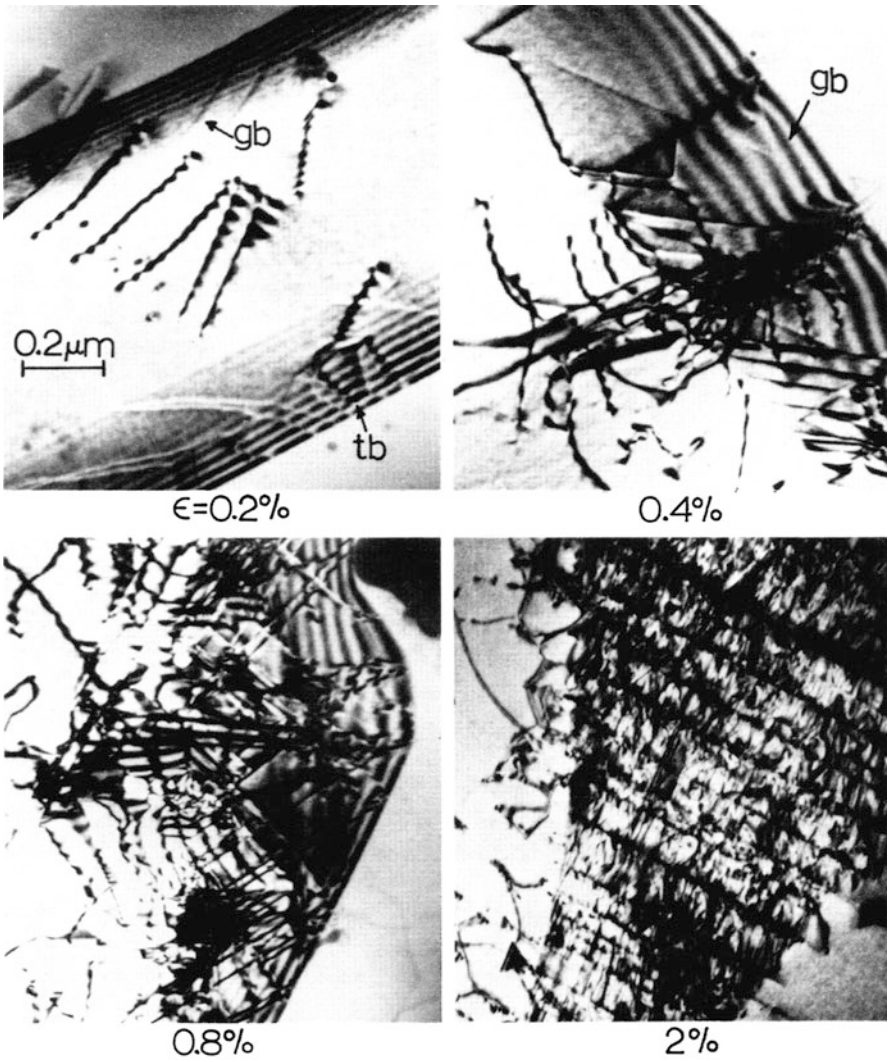


Fig. 31 Changes in grain boundary (ledge) structure with uniaxial strain in 304 stainless steel. Note at $\in = 0.2$ % both the grain boundary (gb) and a coherent twin boundary (tb) contain ledge structure and emit dislocations. Magnifications are as shown for $\in = 0.2$ % and the grain surface orientations are (110) (From Murr and Wang 1982)

- *Pileups of GBDs can occur which lead to the formation of grain boundary ledges* (► Fig. 14a, b of chapter "Line Defects: Dislocations in Crystalline Materials").
- *Since the Burgers vector of a GBD is not restricted to lie in the grain boundary plane, grain boundary sliding may involve both (dislocation) glide and climb.*
- *GBDs may be produced by either grain boundary deformation or by grain boundary migration.*

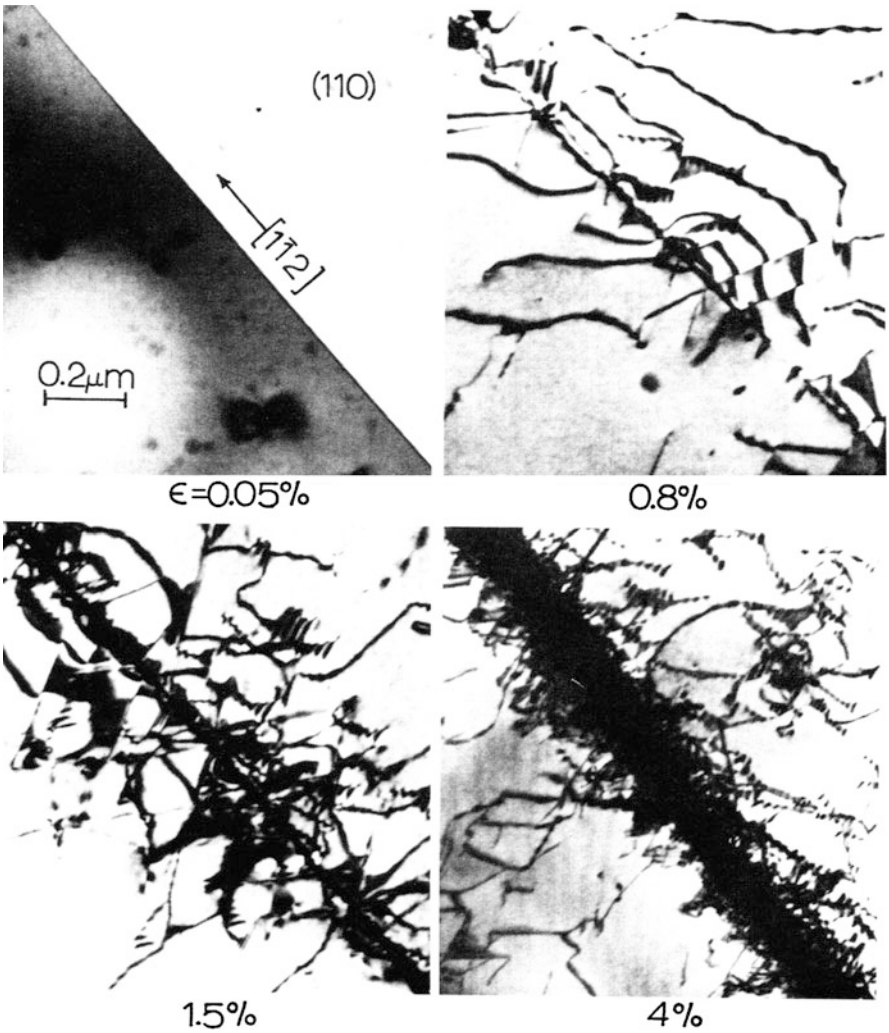


Fig. 32 Systematic changes in dislocation structure adjacent to an identically oriented coherent twin boundary viewed edge on, with changes in uniaxial strain shown. 304 stainless steel sheet. Magnifications are the same as shown for $\in = 0.05$ % (From Murr and Wang 1982)

- *GBDs can cause grain boundary sliding in turn affecting creep and superplastic behavior, create grain boundary ledges, and initiate both tensile and shear cracks leading to both trans- and intergranular fracture.*

Like grain boundaries, phase boundaries and special boundaries, such as semi-coherent interfaces or noncoherent interfaces (which can include phase boundaries), are common solid-state interfaces separating 2-phase regimes having different chemistry and crystal structures (or lattice parameters). Figure 33

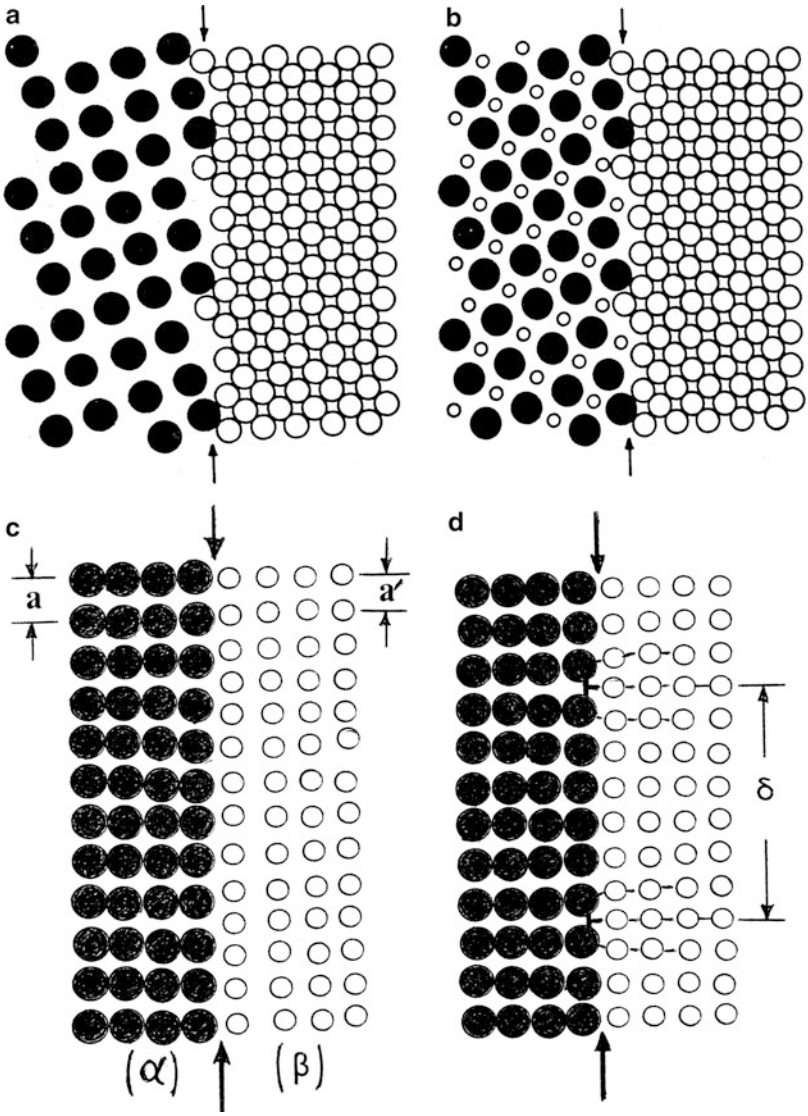


Fig. 33 Interphase interfaces. (**a**) Two single-component phases. Interface is denoted by *arrows*. (**b**) Multicomponent phase boundary (*arrows*). (**c**) Noncoherent interphase boundary separating phase α and β (*arrows*). (**d**) Coherent (semi-coherent) interphase interface with dislocations spaced δ

illustrates these fundamental issues. Figure 33a shows two single-component phase regions, while Fig. 33b shows a binary component phase separated from a single-component phase. Figure 33c shows two different single-component phases with similar crystal structure but different lattice parameters creating a noncoherent

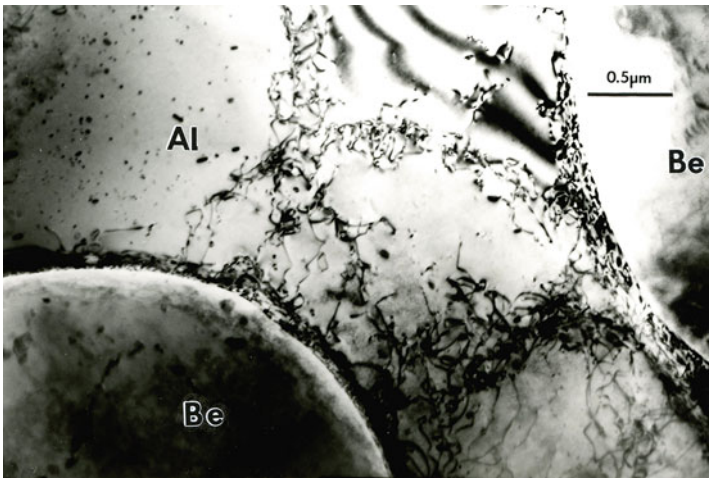


Fig. 34 Be spheres in Al matrix (62 Be-38 Al). AlBeMet AM162 powder metallurgy alloy showing curved interphase boundary

interface. In contrast to Fig. 33c, d shows a semi-coherent interface containing dislocations (often called misfit dislocations) spaced δ. The misfit (as strain) is given by

$$\in_m = \left(\frac{a_\alpha - a_\beta}{a_\alpha}\right) = \left(\frac{a - a'}{a}\right) \times 100\%; \quad = \quad \frac{2\left(a_\propto - a_\beta\right)}{\left(a_\propto - a_\beta\right)} = \frac{2(a - a')}{(a + a')}, \qquad (16)$$

where a and a′ are the lattice parameters corresponding to phase α and β, respectively, in Fig. 33d. The dislocations shown in Fig. 33d have a Burgers vector, b_m, and the corresponding spacing, δ, is given by

$$\delta = \frac{b_m}{\in_m}, \qquad (17)$$

where b_m represents the magnitude of the edge component of the Burgers vector of the misfit dislocation. This semi-coherent boundary in Fig. 33d is similar to the dislocation boundary represented in Fig. 16b, and Eq. 17 is therefore essentially equivalent to Eq. 9 with the misorientation angle ($\Theta \cong 2 \sin(\Theta/2)$) replaced by the misfit, $\in_m$. It can be observed in Fig. 33c that the interphase boundary, along with a rotation normal to the boundary plane, would be represented by Fig. 15g, h, respectively, while this would change with the introduction of the misfit dislocations represented in Fig. 33d.

Figure 34 shows noncoherent interfaces corresponding to either Fig. 33a or c for Be particles in an aluminum matrix, with apparent dislocation emission from the Be/Al phase boundary. While the dislocation emission may be characterized by the ledge structures illustrated for grain boundaries in Fig. 30d and ▶ Fig. 14 of chapter "Line Defects: Dislocations in Crystalline Materials," a similar mechanism has

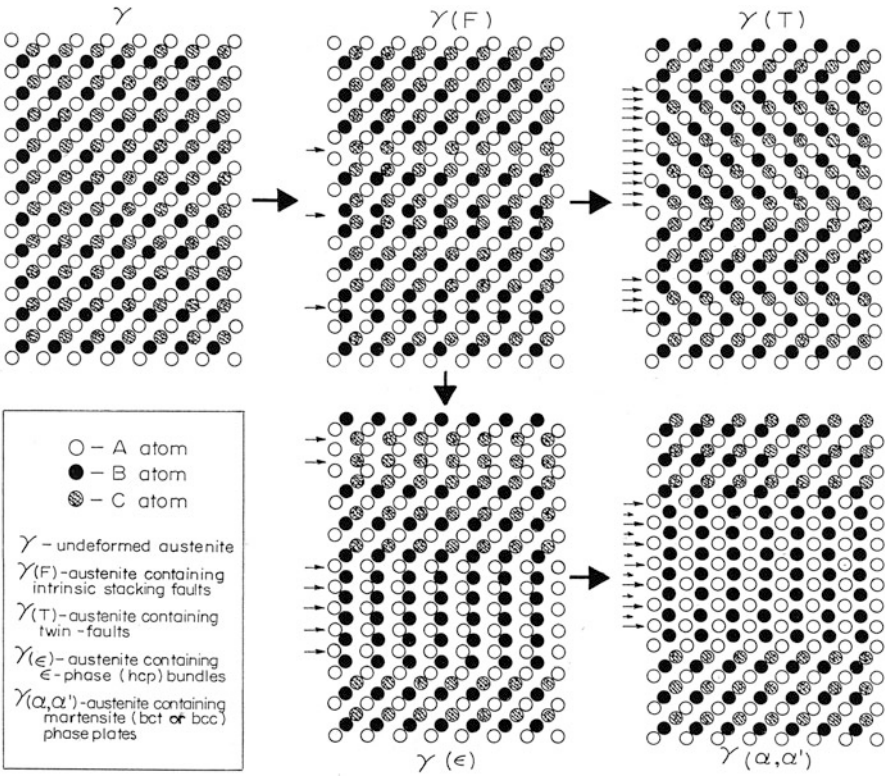


Fig. 35 (110) fcc close-packing schematics illustrating stacking-fault (γ(F)), twin (γ(T)), hcp-epsilon phase (γ(∈)), and bcc/bct phase (γ (α, α′)) formation. *Large arrows* illustrate phase evolution involving $(\bar{1}11)$ plane shift or incremental glide

been discussed for ledge generation of screw dislocations in interphase boundaries by Price and Hirth (1972). Figure 34 also represents the larger features of volume defects represented by Be particles in the Al matrix.

While there are of course many examples of interphase/phase boundaries, phase boundaries created by phase transformations can also be represented by systematic interactions of dislocations or dislocation-related structures such as stacking faults or stacking sequences characterized by strain invariant conditions. Figure 35 illustrates these mechanisms for strain-induced twin, hcp-epsilon (∈) phase and bcc/bct-martensite phase formation in austenite (stainless steels). Figure 36 illustrates these phase formations (or transformations) in deformed type 304 stainless steel. The bct (α′) martensite in Fig. 36 forms by intersecting stacking faults (a <112>/6) and stacking faults with slight displacement variances corresponding to (3/8) (a <112>/6) or a <112>/16, along with displacements of the form a <112>/18 (Staudhammer et al. 1983), forming ε and bct-phase regions illustrated in Fig. 35 (γ (α, α′)), and in the dark-field TEM image in Fig. 36e.

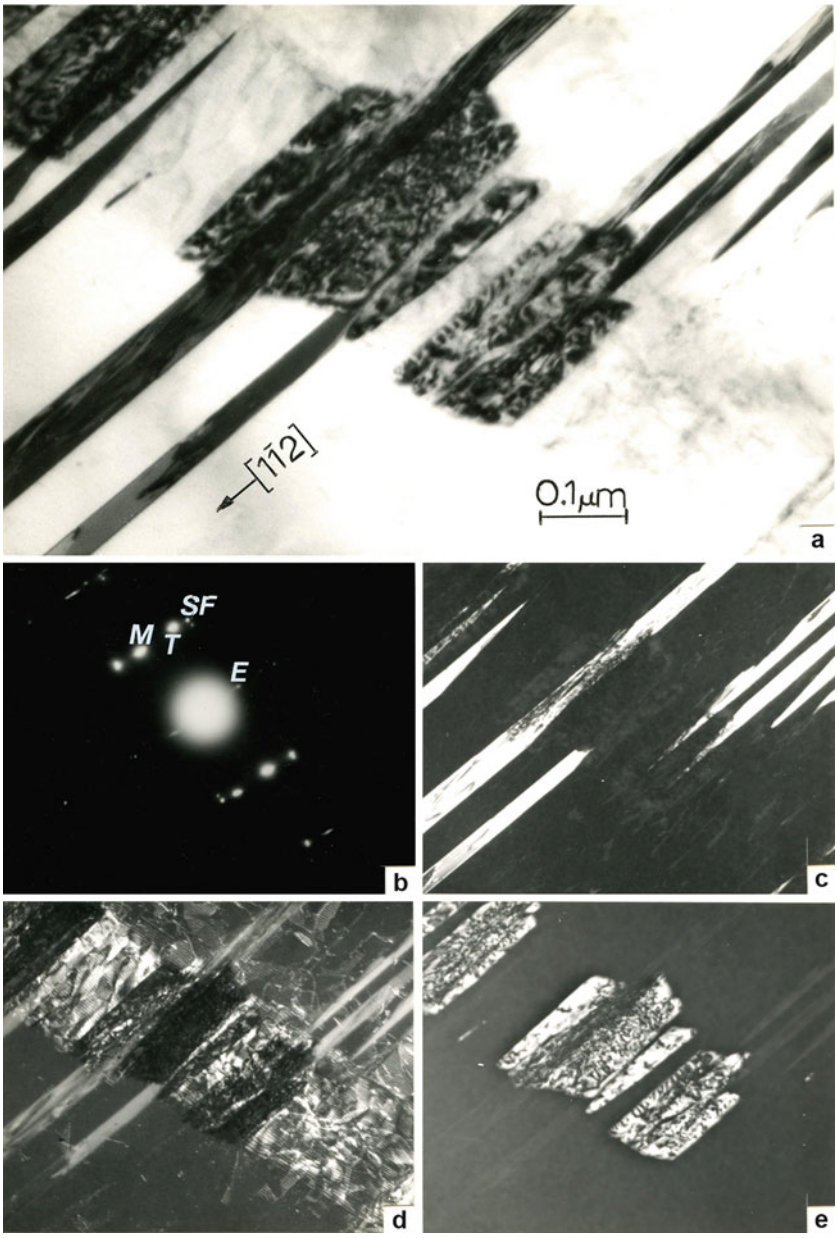


Fig. 36 TEM image sequence showing strain invariant-induced martensite (bct) formation and associated interfacial contributions (as in ▶ Fig. 75 of chapter "A Brief Introduction to Quantum Mechanics"). (**a**) $(\bar{1}11)$ twin faults intersecting martensite phase. (**b**) Selected-area electron diffraction pattern: *M* martensite spot, *T* twin spot, *SF* stacking-fault reflection, $\in$-epsilon spot. (**c**) Dark-field image of twins in (**a**) using T in (**b**). (**d**) Dark-field image for T and M in (**b**). (**e**) Martensite phase using M in (**b**)

References

Bernstein IM, Rath BB, Thomas LE (1972) Studies of fracture in the high voltage electron microscope In: Thomas G, Fulrath RM, Fisher RW (eds) Electron microscopy and the structure of materials. University of California Press, Berkeley, p 75

Cottrell AH (1953) Dislocations and plastic flow in crystals. Clarendon Press, Oxford

Huang JC, Gray GT III (1989) Microband formation in shock-loaded and cold-rolled metals. Acta Metall 37(2):3335–3347

Klemm V, Klimanek P, Motylenko M (2002) Transmission electron microscopy analysis of disclination structures in plastically deformed metals. Mater Sci Eng A 324:174–181

Li JCM (1961) High-angle tilt boundary: a dislocation core model. J Appl Phys 32:525–5301

Luft A (1991) Microstructural processes of plastic instabilities. Prog Mater Sci 35:97–204

Marcinkowski MJ (1972) Dislocation behavior and contrast effects associated with grain boundaries and related internal boundaries. In: Thomas G, Fulrath RM, Fisher RM (eds) Electron microscopy and structure of materials. University of California Press, Berkeley, pp 383–416

Murr LE (1975) Interfacial phenomena in metals and alloys. Addison-Wesley, Reading

Murr LE (1981) Residual microstructure-mechanical property relationships in shock-loaded metals and alloys, Chapter 37. In: Meyers MA, Murr LE (eds) Shock waves and high-strain rate phenomena in metals: concepts and applications. Plenum, New York, pp 607–673

Murr LE (1988) Examination of microstructural development by shock waves in condensed matter: theoretical and practical consequences In: Schmidt SC, Holmes NC (eds) Shock waves in condensed matter. Elsevier, Amsterdam, pp 315–320

Murr LE (1991) Explosive processing of bulk superconductors. Mater Manuf Process 6(1):1–31

Murr LE, Esquivel EV (2004) Review; observations of common microstructural issues associated with dynamic deformation phenomena: twins, microbands, grain size effects, shear sands, and dynamic recrystallization. J Mater Sci 39:1153–1168

Murr LE, Wang S-H (1982) Comparison of microstructural evolution associated with the stress-strain diagram for nickel and 304 stainless steel: an electron microscope study of microyielding and plastic flow. Res Mech 4:237–274

Murr LE, Horylev RJ, Lin WN (1970) Interfacial energy and structure in fcc metals and alloys. Philos Mag 22(177):515–542

Price CW, Hirth JP (1972) A mechanism for the generation of screw dislocations from grain-boundary ledges. Mater Sci Eng 9:15–18

Quinones SA, Murr LE (1998) Correlation of computed simulations with residual hardness mapping and microstructural observations of high velocity and hypervelocity impact craters in copper. Phys Stat Solid 166:763–773

Read WT, Schockley W (1950) Dislocation models of crystal grain boundaries. Phys Rev 78:275–279

Ruff AW, Ives LK (1973) Stacking fault energy in silver-tin alloys. Phys Stat Solid 16:133–139

Staudhammer KP, Murr LE, Hecker SS (1983). Electron microscope study of the genesis of strain-induced martensite embryos. In: Aronson HI (ed) Proceedings of international conference on solid-state transformations. Carnegie-Mellon University, Pittsburgh, pp 1287–1291

Weertman J, Weertman JR (1992) Elementary dislocation theory. Oxford University Press, Oxford

Volume Defects: 3D Imperfections in Crystals

Contents

Abstract

Second-phase (solid-state) inclusions such as precipitates and dispersed phase particles are discussed briefly along with other clusters, voids, and bubbles, such as those which can arise by radiation effects. Other types of defects involving carbon aggregates forming various graphite inclusions are discussed in later chapters dealing with such features in iron and related carbides in steels.

Introduction

The aggregation of point defects, such as vacancies, and interstitial elements, such as carbon, creates a range of volume inclusions or defects which can range from the small nano-regime (~1–5 nm) to the micron regime. Vacancy aggregation can form various void volumes, while neutron and other radiation effects can produce gas-filled (He) bubbles especially in metals and alloys, which can have dramatic effects on the properties and performance of materials. In addition, precipitation or dispersed particle aggregates can form in similar ways as directed by diffusional phenomena described previously in chapter "▶ Point Defects." Solute concentration as it is influenced by temperature and time is implicit in Fick's first law (▶ Eq. 4 of chapter "Point Defects").

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_17

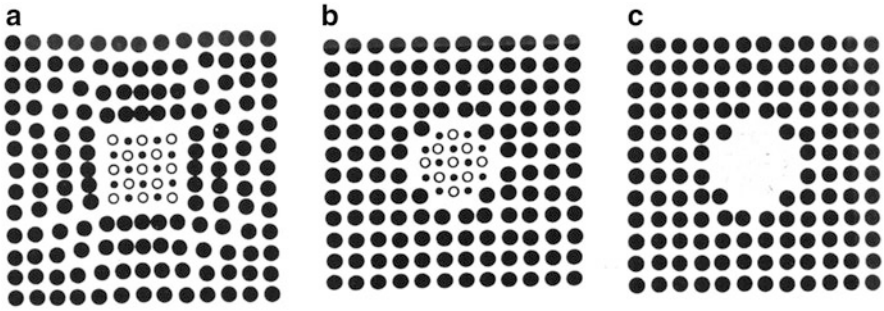


Fig. 1 Volume defects. (**a**) Coherent phase volume. (**b**) Noncoherent phase volume. (**c**) Void (faceted)

Examples of Volume Defects: Second-Phase Solid-State Inclusions and Free Volumes (Voids and Bubbles)

Volume defects in crystalline material can constitute small clusters of atoms such as copper in aluminum alloys forming *Guinier-Preston zones* or metal aggregates formed from F-centers in NaCl or CaF_2 illustrated in ▶ Figs. 10 and ▶ 11 (chapter "Electromagnetic Color and Color in Materials"), second-phase particles or precipitates, and voids formed from vacancy aggregation or gas atoms, such as He, during high-energy irradiation. These 3D clusters or second-phase particles can vary in size from less than a nanometer to more than a centimeter. ▶ Figure 34 of chapter "Planar Defects: Crystal Interfaces" shows micron-size Be particles in an Al matrix, while ▶ Fig. 36 of chapter "Planar Defects: Crystal Interfaces" shows submicron-size martensite forming in austenite. In the case of the micron-size Be particles in the Al matrix of ▶ Fig. 34 of chapter "Planar Defects: Crystal Interfaces," the imperfection may in fact reside in the interphase boundary separating the Be from the Al and characteristic of the interface illustrated schematically in ▶ Fig. 33a of chapter "Planar Defects: Crystal Interfaces."

Figure 1 illustrates coherent and noncoherent precipitates or second-phase particles in a crystal matrix (▶ Fig. 37a, b of chapter "Planar Defects: Crystal Interfaces"), along with a faceted void region resulting by vacancy migration (diffusion) and clustering, or gas atom diffusion and clustering. Figure 1a, b corresponds, from an interfacial or phase boundary perspective, to ▶ Fig. 33d and c of chapter "Planar Defects: Crystal Interfaces." Figure 2 shows examples of noncoherent and coherent inclusions which correspond in principle to the schematic views shown Fig. 1b, a, respectively. The ThO_2 particles dispersed in a Ni matrix in Fig. 2a are noncoherent, and the insert (lower left) shows diffraction contrast fringes within a larger particle indicating the particle, and others are in fact single-crystal ThO_2 (with the cubic fluorite (CaF_2) structure (▶ Fig. 11a of chapter "Electromagnetic Color and Color in Materials")). In Fig. 2b, very tiny (<10 nm)

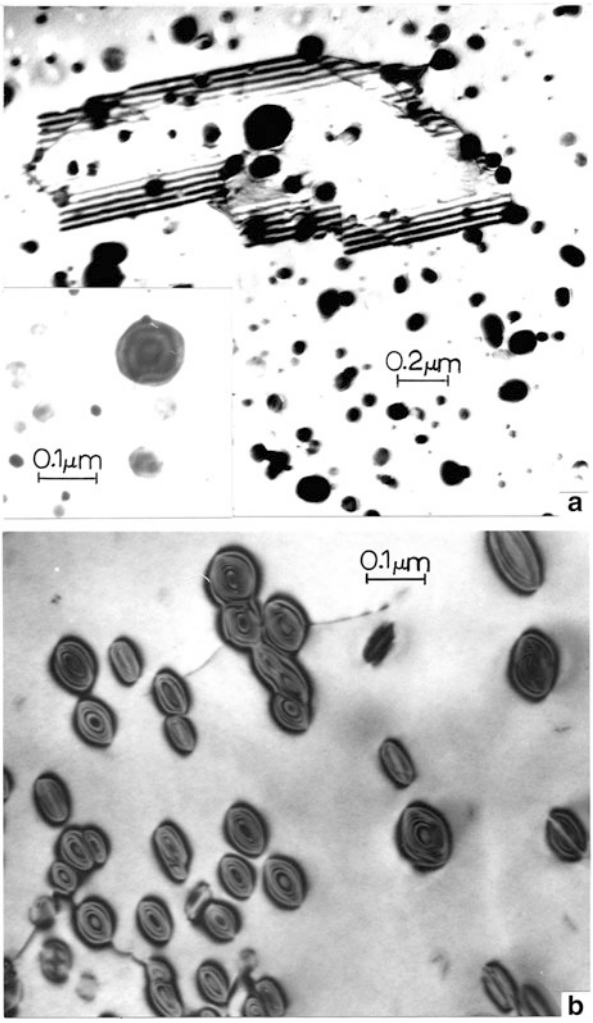


Fig. 2 (**a**) Noncoherent (ThO_2) dispersed (single-crystal) particles in Ni. Occluded twin with particles on noncoherent boundaries. *Insert* shows diffraction contrast in particle. (**b**) Coherent Cr_6C precipitates showing strain-field diffraction contrast surrounding particles. (**a**) and (**b**) are TEM bright-field images

Cr_6C precipitates create very large coherency strain-field contrast surrounding them, observed by TEM imaging.

Aluminum-copper alloys such as Al-606 l (Al-4 % Cu) provide some interesting cases for coherent-noncoherent precipitate phenomena, depending upon aging times and temperature as illustrated in Fig. 3. In Fig. 2, so-called Guinier-Preston (GP1) zones form as copper-rich clusters replacing Al atoms in the {100} planes of the fcc alloy structure. Since the Cu atoms are smaller than the Al atoms, these regions create large strain fields parallel to the {100} planes which they occupy, and these are clearly visible in Fig. 3b. By tilting the area in the TEM, the thin cluster zones become very weak, thin lines as shown in Fig. 3c, while with slightly different aging, these precipitates become noncoherent and assume a composition

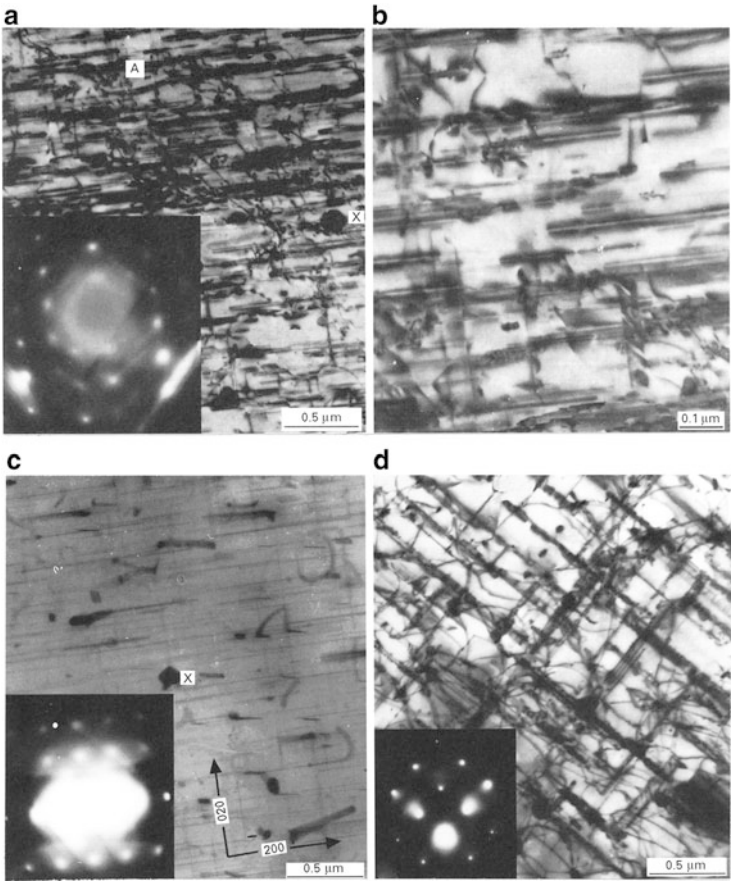


Fig. 3 Bright-field TEM images showing {100} coincident coherent and noncoherent precipitates in Al-6061 alloy. (**a**) (100) precipitates. Diffraction pattern insert is (100). (**b**) Magnified and tilted view at *A* in (**a**) showing linear coherency strain-field contrast for Cu sheets in (010)-GP1 zones. (**c**) Tilted image of (**a**) at *x* showing fine precipitate planes in (100) and (010). Diffraction pattern insert is (100). (**d**) Noncoherent precipitates in (100) and (010) planes. Diffraction pattern insert is (100) (Adapted from Murr et al. (1998))

of Al_2Cu and are illustrated in Fig. 3d. The coherent GP1 zone structure in Fig. 3b evolves into the noncoherent, platelike GP2 (or θ'') precipitate structure in Fig. 3d.

Different aging environments continue to create an evolution of precipitation as shown in Fig. 4 in contrast to Fig. 3. In Fig. 4a, thin platelike precipitates evolve as θ' (tetragonal Al_2Cu: a = 4.04 Å, c = 5.8 Å) along with very tiny precipitates which may be nucleating θ (also tetragonal with a = 6.06 Å, c = 4.87 Å). These precipitates are noncoherent, and a close examination (although difficult from Fig. 4a) reveals apparent dislocation emission from small interfacial ledges. In Fig. 4b, needlelike precipitates appear crystallographically as in Fig. 4a, along with larger, presumably θ precipitates, both of which are noncoherent, but

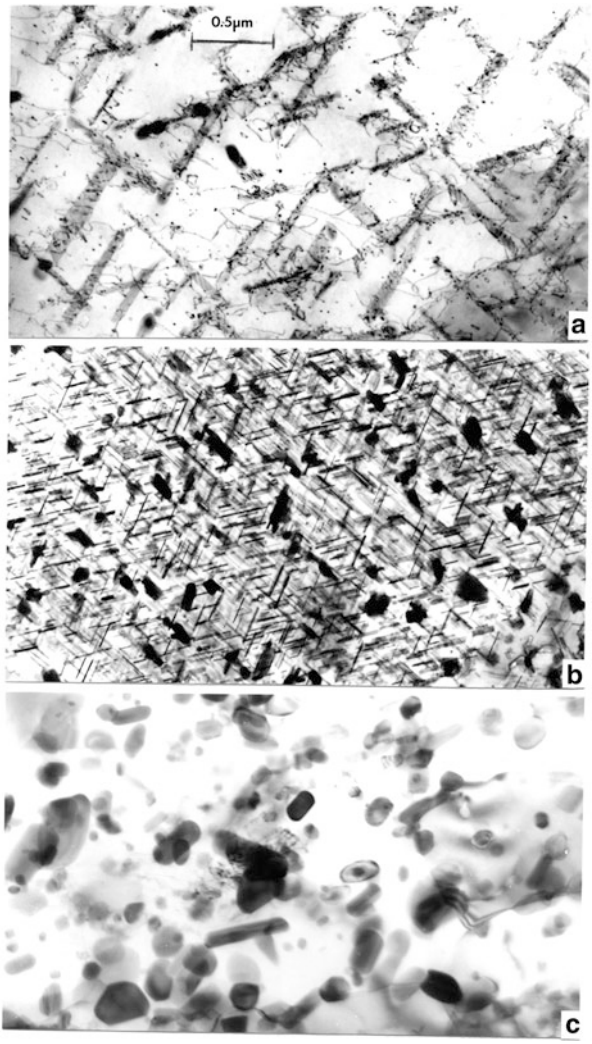


Fig. 4 Contrasting noncoherent precipitate morphologies in Al-6061 alloy. (**a**) Precipitate plates in {110} planes along with very tiny precipitate particles, dislocation loops, and other dislocation structures. (**b**) Needlelike precipitates in {110} planes forming Widmanstätten patterns along with other precipitate morphologies. (**c**) Random precipitate variety showing various shapes and facets. Magnification is the same for all images and is given in (**a**)

coincident with the matrix {110} planes. These very closely spaced needle precipitates are often referred to as a Widmanstätten structure, named after Alois von Beckh Widmanstätten who, in 1808, noticed color- and luster-zone patterns on heating iron meteorites, although G. Thompson published similar findings for nitric acid-treated meteorites 4 years earlier. These closely spaced precipitates create finely partitioned grain structures and noticeably increase strength and hardness. Finally, as illustrated in Fig. 4c, long-time aging at higher temperatures can produce a plethora of large, often noticeably faceted, noncoherent precipitates with a variety of shapes and sizes, as well as stoichiometries differing from Al_2Cu. These precipitates are a smaller version of the second-phase Be inclusions in an aluminum

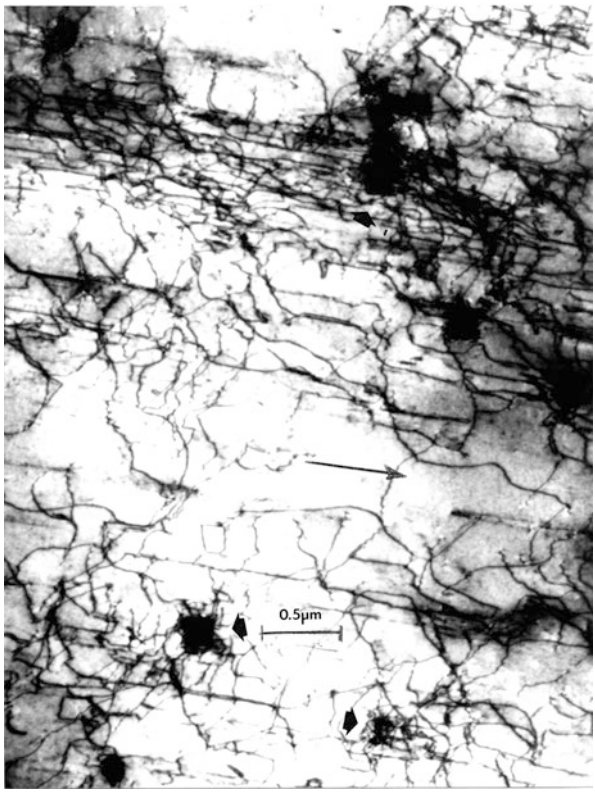


Fig. 5 Cr_6C precipitates in Inconel 600 (▶ Fig. 78b of chapter "A Brief Introduction to Quantum Mechanics") become noncoherent and emit dislocations which surround them after plane wave shock loading at 18 GPa (*Large arrows*). Smaller (long) *arrow* shows $[1\bar{1}0]$ trace direction for the (110) grain surface orientation

matrix shown in ▶ Fig. 34 (chapter "Planar Defects: Crystal Interfaces"), but are in fact similar to the noncoherent ThO_2 single-crystal dispersed particles shown in Fig. 2a. All represent 3D imperfections separated by interphase boundaries from the matrix in which they occur.

The accommodation of very single-crystal small precipitates can involve small steps or ledge structures which can become active sources in the production of dislocations. Even when coherent precipitates undergo severe deformation, the coherent or semicoherent interface may become noncoherent and activate ledge sources of dislocations. This phenomenon is illustrated in Fig. 5 which shows the same alloy containing small, coherent Cr_6C precipitates illustrated in Fig. 2b which become noncoherent and produce a high density of dislocations at their interphase boundary after plane shock wave passage. In these cases, the precipitates become interior grain sources for dislocations and increase the overall strength accordingly.

The ability to systematically and selectively develop a variety of precipitates in aluminum alloys as illustrated in Figs. 3 and 4 points up the unique features of these alloys and especially Al-6061 which has at least six heat-treatment (T) schedules in addition to the fully annealed material referred to as T(0). In this regard, the ability to systematically develop precipitates and manipulate their volume fraction, and

their degree of coherency (or misfit as represented in ▶ Eq. 16 of chapter "Planar Defects: Crystal Interfaces"), has been instrumental in developing modern-day aircraft and related gas-turbine engine valves and blades from nickel-based alloys strengthened by the Ni_3 (Al, Ti) γ' precipitate phase. In 1941, the introduction of Nimonic 80 alloy and later (in 1945) the progressively stronger Nimonic 90 alloy (54Ni – 18 to 21Cr – 15 to 21Co – 2 to 3 Ti – 1 to 2 Al; in weight percent) became the precursors of modern Ni-based superalloys which have included Inconels, MAR-M, Haynes, UDIMET, and Rene alloy systems with controlled additions of W, Ta, Mo, and Nb to solution-harden the γ (Ni-Cr fcc matrix) phase, along with additions of Al, Ti, or Ta to precipitate the γ' phase which is coherent and coplanar with the fcc γ matrix. C, B, and Zr have also been added to promote ductility, and depending upon heat treatment, carbon additions can react to form various carbide phases (MC, $M_{23}C_6$, M_6C (as in Figs. 2b and 5), and M_7C_3, where M is usually Cr). The volume fraction (and coarsening kinetics) of the γ' phase precipitates in Ni-based superalloys depends not only on the Al and Ti content but also on the additions of Nb, Ta, W, and Mo in the alloy composition, since these elements can substitute for Al atoms in the Ni_3Al fcc structure, represented ideally by the Ll_2-type (fcc)-ordered structure in ▶ Fig. 1b (chapter "Structure of Metals and Alloys"). Additionally, the γ' precipitate morphology is related to the volume fraction of γ' as well as the degree of γ/γ' misfit (▶ Eq. 16 of chapter "Planar Defects: Crystal Interfaces"), with cuboidal precipitate particles representing the more preferred morphology and characteristic of some degree of misfit. In addition to compositional tuning in Ni-based superalloys, the elimination of grain boundaries transverses to the stress direction (or normal to the turbine-blade axis) by directional solidification improved ductility and creep rupture life, as well as thermal shock resistance.

Figure 6 illustrates the development of a Rene' 142 (General Electric) alloy (Ni–12Co–6.8Cr–6.3Ta–6.1Al–4.9 W–0.7Mo, balance Ni in weight percent) product by additive manufacturing from pre-alloyed, precursor powder by electron beam melting. Figure 6a shows γ' (Ni_3Al) precipitates with dislocations in the surrounding γ matrix, while Fig. 6b shows dense, well-defined cuboidal γ' precipitates exhibiting systematic δ-fringe contrast arising at the coherent γ/γ' interfaces (Oblak and Kear 1972). Since the γ' has a larger lattice parameter than the fcc γ matrix ($a = 3.58$ Å), a tetragonal distortion of the γ matrix at the γ/γ' interface occurs, and as a consequence, the diffraction vector, **g**, shown in the selected area electron diffraction pattern insert in Fig. 6b, changes discontinuously at the interface. The discontinuity in **g** is perpendicular to the interface which exhibits the strong fringe contrast shown between the γ' precipitate and the γ matrix in Fig. 6b. These features will be explored in greater detail in chapter "▶ Examples of Directional Crystal Structures: Gas Turbine Component Applications in Superalloys."

There are numerous examples of void or void-related phenomena in variously irradiated crystalline materials. The formation of various types of color centers, particularly F-centers as illustrated for CaF_2 in ▶ Figs. 11 and ▶ 12 (chapter "Electromagnetic Color and Color in Materials"), is typical. Natural CaF_2 as shown in ▶ Fig. 12 of chapter "Electromagnetic Color and Color in Materials" is

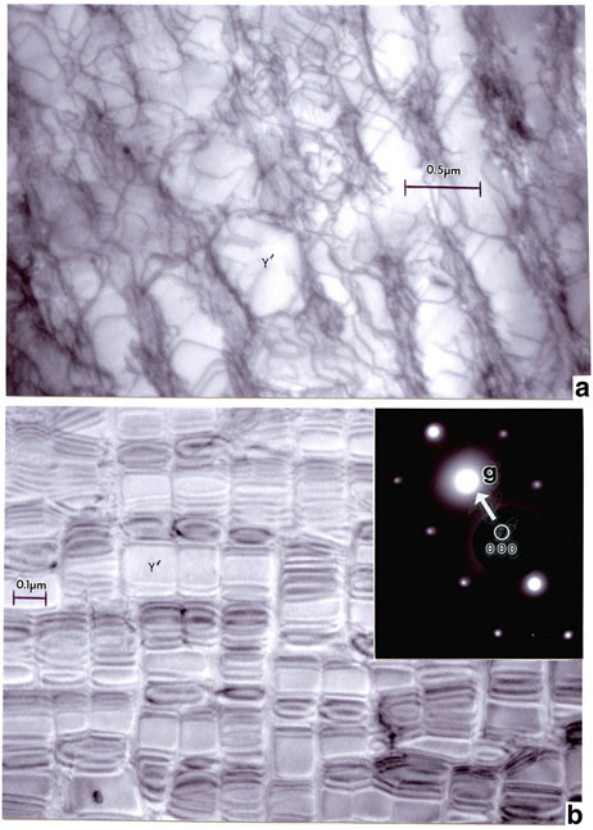


Fig. 6 TEM images for cuboidal γ' (Ni_3Al) precipitates in Rene'142 nickel-based superalloy (fcc (γ) Ni-Cr solid solution matrix). (**a**) Tilted for dislocation structure in γ region separating γ'. (**b**) δ-fringe contrast arising by coherency strain fields at γ/γ' coherent interface. Operating reflection $\mathbf{g} = [1\bar{1}1]$. Diffraction pattern insert shows (110) orientation

irradiated by radioactive inclusions such as uranium or thorium. Uranium 238 produces alpha particles which can then form radioactive thorium (Th-234). This decay product can become a beta particle producer. These processes can create F-centers, and over time these can agglomerate to form F-center clusters. Such clusters can appear as solid-state calcium colloids when the electrons occupying F sites combine with Ca^{2+} to form neutral Ca°. Similar phenomena can of course occur in NaCl to produce Na° colloids. However, this process is easily facilitated by irradiation in the transmission electron microscope, and such metal clusters appear during observations of CaF_2 in the TEM as shown in Fig. 7a. These clusters or F-center aggregates form hexagonal arrays as shown in Fig. 7b by the systematic formation of F-centers in the CaF_2 {111} planes illustrated in Fig. 7c–f (Murr 1974).

Electron irradiation has also been observed to form vacancy and/or interstitial loops and stacking-fault tetrahedra in metals and alloys during TEM observation, especially at high electron accelerating voltages while simultaneously heating the material. Vacancy or intrinsic loops (or prismatic loops) in fcc materials (also called *S-Frank loops* (Weertman and Weertman 1992)) in contrast to interstitial or extrinsic loops in fcc materials (also called *D-Frank loops* (Weertman and

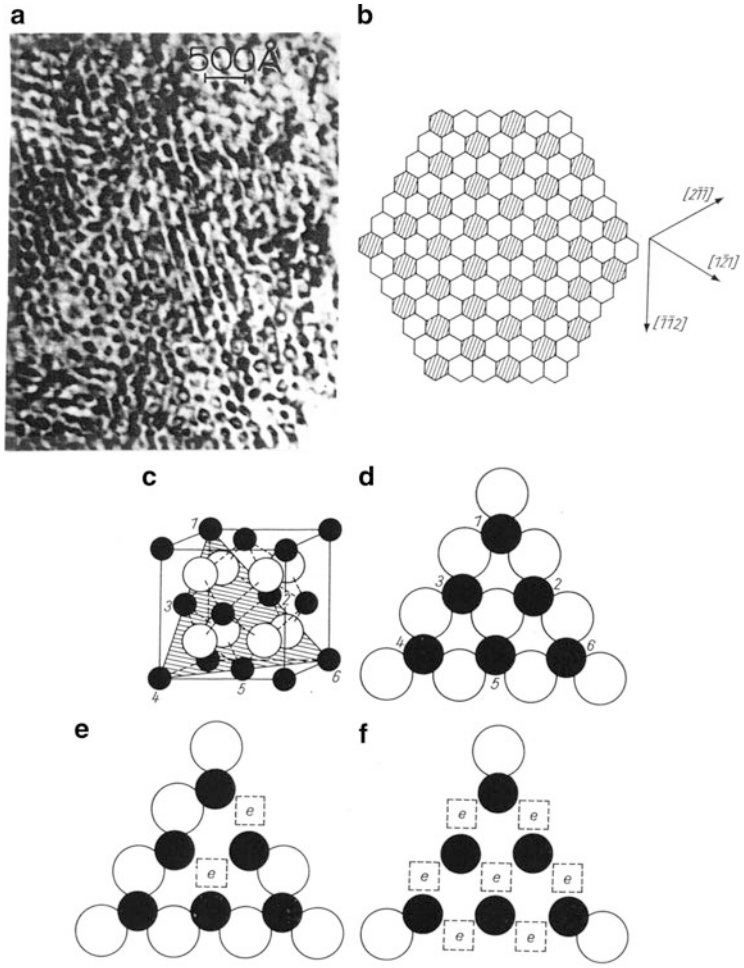


Fig. 7 F-aggregate center formation in fluorite (CaF_2). (**a**) TEM image formed after electron irradiation at 200 kV. (**b**) Schematic representation of F-aggregate centers in CaF_2 (111) planes. (**c**) CaF_2 unit cell highlighting (111) plane. (**d**) Ca^{2+} solid circles, F^- open circles in (111) plane. (**e**) F-centers in (111) plane. (**f**) Hexagonal arrangement of F-center aggregate (Adapted from Murr (1974))

Weertman 1992)) are illustrated schematically in Fig. 8a, b, respectively. With varying radiation energy or fluence and temperature, these loops can expand or form stacking-fault tetrahedra (bounded by total dislocations on the {111} planes corresponding to the Thompson tetrahedron in ▶ Fig. 31 of chapter "A Brief Introduction to Quantum Mechanics") as well as voids in the case of vacancies. Such tetrahedra or voids can also grow with time or temperature as well as the energy of radiation or dose. Figure 8c, d illustrates extrinsic (D-Frank) loops or interstitial loops in electron-irradiated copper and nickel, respectively, where the

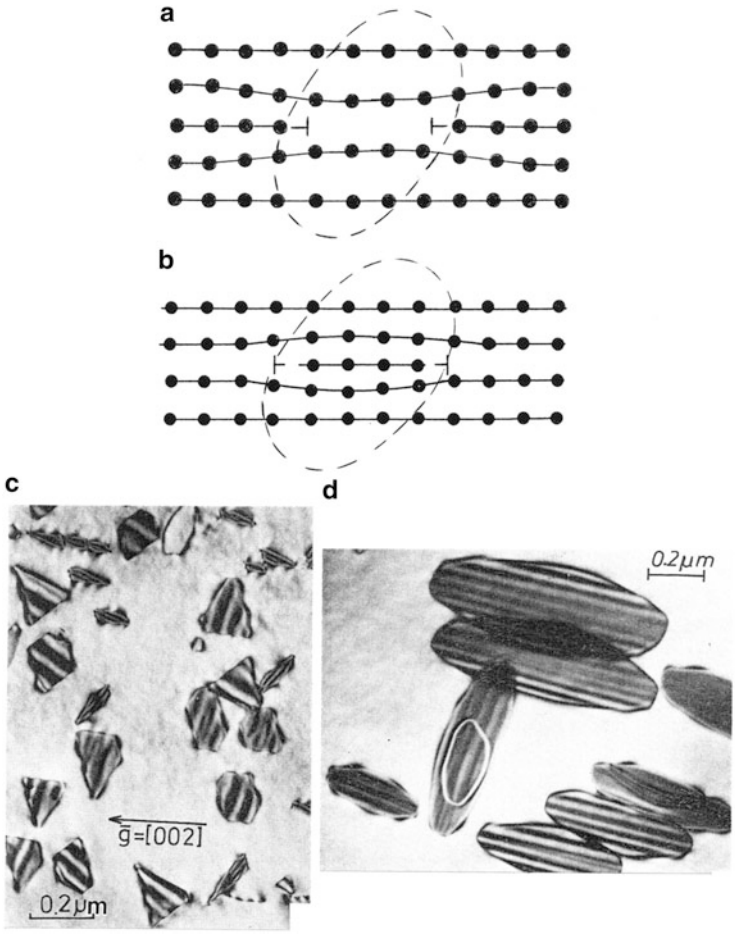


Fig. 8 Frank loops (prismatic dislocation loops) in fcc structures. (**a**) S-Frank (intrinsic) vacancy loop. (**b**) D-Frank (extrinsic) interstitial loop. (**c**) Interstitial (stacking-fault) loops in Cu electron-irradiated (1 MV) 200 min at 200 °C (From Meshii and Shiraishi (1972)). (**d**) Interstitial (stacking-fault) loops in electron-irradiated (1 MV) Ni, nucleated at 450 °C and grown at 570 °C (Adapted from Urban and Wilkens (1972))

threshold voltages for interstitial-vacancy production are correspondingly 400 and 650 kV. These large interstitial loops (Fig. 8c, d) exhibit stacking-fault contrast implicit in the schematic of Fig. 8b. Note the faulted loops are on (111) planes inclined 54° to the (100) grain surface plane and intersect one another at 90° as a consequence of the {111} plane intersections.

Zinkle and Farrell (1989) observed stacking-fault tetrahedra in neutron irradiated copper at 182 °C, but voids formed between 220 °C and 450 °C. In this range, the copper also swelled. Helium bubbles in Ni have also been observed to form faceted voids, with a preference for void formation along grain boundaries

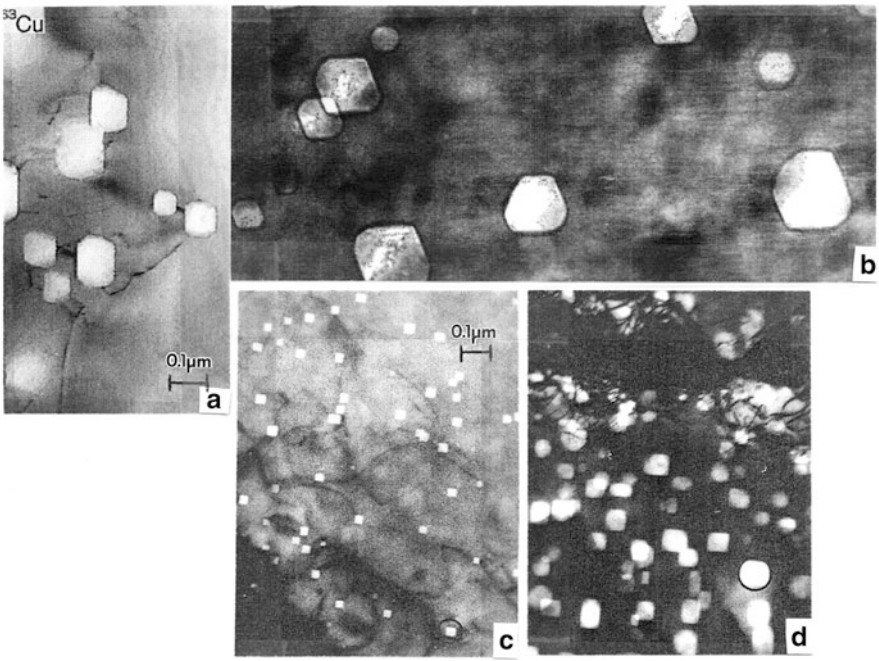


Fig. 9 Voids produced by electron irradiation. (**a**) and (**b**) show voids in Cu. Note faceting similar to precipitates of essentially the same size in ▶ Fig. 80b–d of chapter "A Brief Introduction to Quantum Mechanics" shows void growth in electron-irradiated Ni at 440 °C for 180 and 667 min., respectively (adapted from Urban and Wilkens (1972). (**b**) is adapted from Wolfer (2000))

(Chernikov et al. 1997). Figure 9 illustrates some examples of void formation and void faceting in irradiated Cu and Ni.

Radiation-induced void formation and especially helium void swelling is particularly important in plutonium aging effects because, as discussed earlier (chapter "▶ Structure of Metals and Alloys"), it decays to uranium by alpha-particle emission (an energetic helium nucleus He^{++}). Damage is also caused by the energetic uranium nuclei which knock Pu atoms from their lattice sites, creating vacancies. Consequently, He in combination with lattice vacancies creates voids (as in Fig. 9) and void swelling. This self-radiation over time causes dimensional changes and compromises the strength of Pu in weapons (▶ Fig. 13 of chapter "Structure of Metals and Alloys"). Voids or bubbles are in equilibrium when the helium gas pressure (P) inside the bubble equals the surface tension or specific surface free energy, F_s:

$$P = 2F_s/r, \tag{1}$$

where r is the bubble radius. Since F_s will differ for different crystal orientations, bubbles and voids tend to facet as illustrated in Fig. 9. As helium accumulates and

bubbles grow, swelling will increase linearly in proportion to the helium concentration (Wolfer 2000).

It is also important to note that radiation can also induce a δ-Pu-to-α-Pu transformation especially in marginally stabilized Pu alloys which could result in significant shrinkage because, as illustrated in ▶ Fig. 9 of chapter "Structure of Metals and Alloys," the α phase has a specific volume ~20 % less than the δ-phase. Consequently, void or bubble swelling may be somewhat offsetting, although the phase change may adversely affect performance, especially detonation response and achievement of critical mass (▶ Fig. 13 of chapter "Structure of Metals and Alloys").

References

Chernikov VH, Trinkaus H, Ullmaier H (1997) Helium bubbles in nickel arranged at $T > 0.7\ T_m$. J Nucl Mater 250:103–110

Meshii M, Shiraishi K (1972) In: Thomas G, Fulrath RM, Fisher RM (eds) Electron-irradiation effects on the structure of copper and aluminum. University of California Press, Berkeley, p 952

Murr LE (1974) Transmission electron microscope study of crystal defects in natural fluorite. Phys State Solid 22:239–256

Murr LE, Liu G, McClure JC (1998) A TEM study of precipitation and related microstructures in friction-stir welded 6061 aluminum. J Mater Sci 33:1243–1251

Oblak JM, Kear BH (1972) Analysis of microstructures in nickel-base alloys: implications for strength and alloy design. In: Thomas G, Fulrath RM, Fisher RM (eds) Electron microscopy and structure of materials. University of California Press, Berkeley, pp 566–616

Urban K, Wilkens (1972) In: Thomas G, Fulrath RM, Fisher RM (eds) Electron microscopy and structure of materials. University of California Press, Berkeley, p 935

Weertman J, Weertman JR (1992) Elementary dislocation theory. Oxford University Press, Oxford, OK

Wolfer WG (2000) Radiation effects in plutonium. Los Alomos Sci 26:275–285

Zinkle SJ, Farrell K (1989) Void swelling and defect cluster formation in reactor-irradiated copper. J Nucl Mater 168:262–267

Strengthening by Crystal Imperfections

Contents

Abstract

Crystal imperfections affect the stress–strain diagram in a number of ways, often increasing the tensile strength as the density of imperfections increase or their geometry creates partitioning of the crystal structure. Work hardening and strain hardening are also dependent upon imperfections. Hardness and yield stress for many metals and alloys are related by a simple relationship, $\sigma_y = H/3$, while the yield stress is often expressed by a Hall–Petch-type relationship, dependent upon the reciprocal square root of the grain size. These issues are presented in this chapter.

Introduction

Imperfections influence the production, interaction, and multiplication of dislocations, including other dislocations. These effects are represented in the form of the plastic region characteristic of the stress–strain diagram. The form of the plastic region responds to work hardening or strain hardening which can be a measure of the immobility of dislocations. While the stress–strain diagram may provide a rough engineering assessment of the plastic behavior of a material, it corresponds to a fixed temperature, usually room temperature, and a low strain rate ($\sim 10^{-3}$/s). In addition, it corresponds to a uniaxial stress or strain. Consequently, it is important to at least examine the stress–strain–strain rate, temperature, and strain-state (uniaxial, biaxial, triaxial) behavior of deformed or

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_18

deforming materials and the role crystal imperfections will have. For polycrystalline materials, the grain size becomes the upper bound for constraint, while this constraint is altered by the incorporation of imperfections: point defects, dislocations, planar defects (stacking faults, twins, etc.), and a host of second phases or volume defects whose dimensions and spacings within the grain can alter the tensile and related stress, hardness, and other mechanical properties, electrical properties, and the like.

Work Hardening or Strain Hardening

Strengthening of materials, in particular crystalline (or polycrystalline) materials, is already implicit in discussions of point defects and planar defects, as well as precipitates and dispersed particles in the context of volume defects. All of these phenomena create specific microstructures in contrast to perfect single crystals, and these microstructures are primarily responsible for predictable properties which influence a material performance.

Historically, and fundamentally as implicit in tensile stress–strain behavior, especially of metals and alloys, strengthening occurs by so-called strain hardening or work hardening according to ▶ Eq. 11 in chapter "Line Defects: Dislocations in Crystalline Materials". In the context of the stress–strain diagram (▶ Fig. 7 of chapter "Line Defects: Dislocations in Crystalline Materials" and Fig. 1), this is often represented for polycrystalline metals as

$$\sigma = \sigma_0 + K\in^n, \tag{1}$$

or

$$\sigma = K\in^n, \tag{2}$$

where K is a material's constant and n is often referred to as the work-hardening coefficient (dependent upon the material – and such features as stacking-fault free energy – temperature of deformation, and the strain, $\in$). Values of n normally vary from ~0.2 to 0.5. In Eq. 1, $\in$ is the true plastic strain while in Eq. 2, the true total strain. The coefficient, K, in Eq. 2 can vary from G/100 to G/1,000, where G is the shear modulus. In addition, strain rate ($\dot{\in}$) can influence the shape of the stress–strain diagram because, in extreme circumstances, heating, which can be localized and adiabatic, is given generally by

$$\Delta T \propto \in \cdot \dot{\in}. \tag{3}$$

Additionally, strain state can have a notable influence on microstructure development, where generally we might indicate that

$$\sigma_i = f(\in, \dot{\in}, T)_i, \tag{4}$$

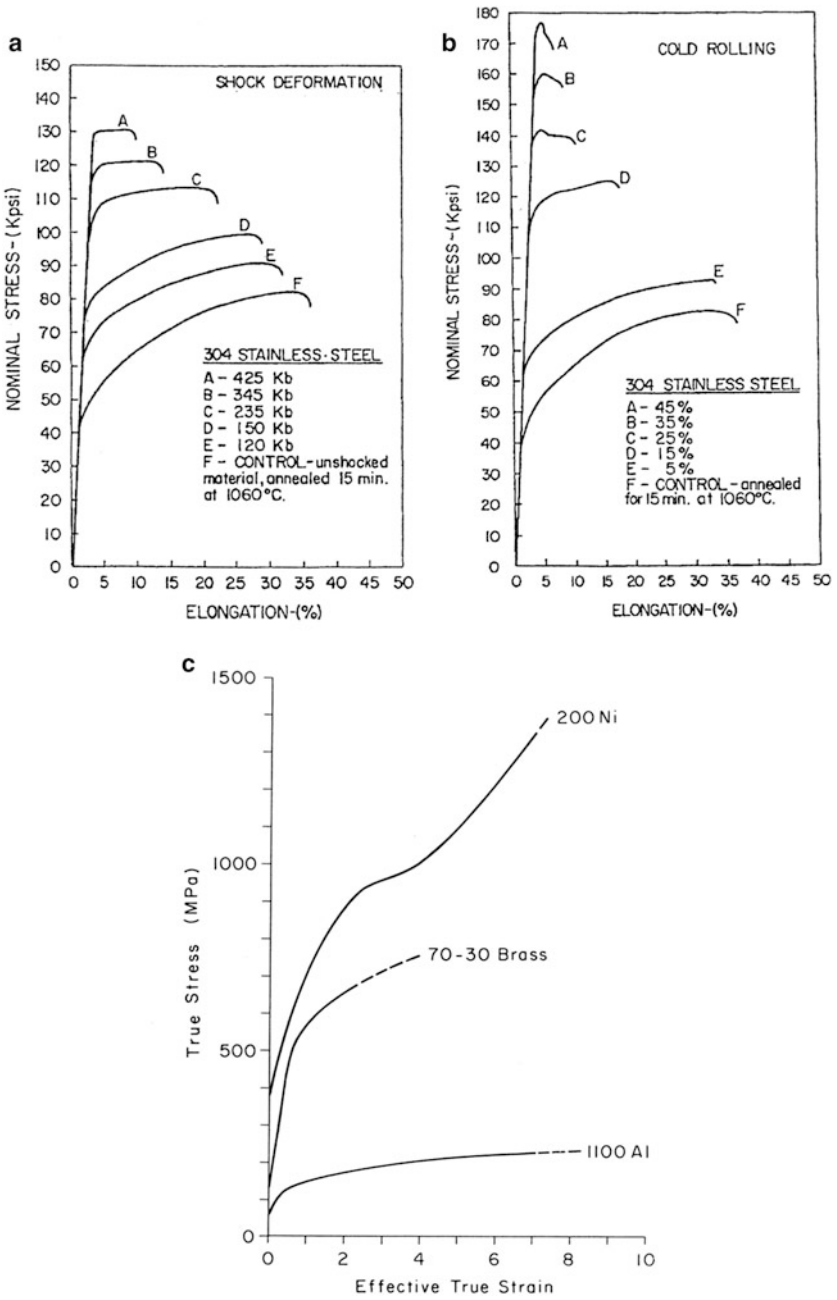


Fig. 1 Tensile straining of variously deformed metals and alloys. (a) 304 stainless steel deformed by uniaxial (plane) shock and tensile tested at 10^{-3}/s. (b) 304 stainless steel cold-rolled and uniaxially tested at 10^{-3}/s. (**c**) True stress–true strain curves for cold-rolled fcc metals

where i represents the three strain or stress states: i = 1 (uniaxial), i = 2 (biaxial), and i = 3 (triaxial). Changes in stress can be expressed as

$$\left[d\sigma = \left(\frac{\partial \sigma}{\partial \epsilon}\right)_{\dot{\epsilon},T} + \left(\frac{\partial \sigma}{\partial \epsilon}\right)_{\dot{\epsilon},T} + \left(\frac{\partial \sigma}{\partial \epsilon}\right)_{\dot{\epsilon},\dot{\epsilon},} \right]; \tag{5}$$

corresponding to stress states, i = 1, 2, and 3: uniaxial, biaxial, and triaxial, respectively.

There is little, comprehensive data comparing $\sigma_{,} \in, \dot{\epsilon}$ at a fixed temperature, let alone comparative temperatures, as well as different stacking-fault free energies or grain sizes, etc. for metals and alloys. Some sense of strengthening and deformation response must therefore be gleaned from a perusal of what might be regarded as a collage of stress, strain, and strain rate performance characteristics for some representative metals and alloys. This begins with some comparative stress–strain (elongation) diagrams for a few common metals and alloys ranging in stacking-fault free energy from ~15 mJ/m^2 for brass to ~180 mJ/m^2 for aluminum (see also ▶ Table 2 of chapter "Planar Defects: Crystal Interfaces") (Murr 1975). However, Fig. 1c does not indicate any systematic trend for stacking-fault free energy. Figure 1a, b shows significant differences in the overall signature of the stress–strain diagram for stainless steel deformed uniaxial (compressive) shock loading ($\dot{\epsilon} \sim 10^6$/s), in contrast to cold-rolled stainless steel. In each case, the samples were shocked or cold-rolled and then deformed by uniaxial, tensile straining at $\dot{\epsilon} \sim 10^{-3}$/s. In Fig. 1c, the materials have been cold-rolled prior to tensile straining and, as illustrated in Fig. 2, contain some significant dislocation density which depends on the degree of rolling deformation implicit on comparing Fig. 2a, b.

Figure 3a compares mild steel deformation as a function of stress, strain, and strain rate at room temperature where increasing both strain and strain rate increases the corresponding stress. Figure 3b shows the production of α′-martensite fraction to be influenced by strain, strain rate, and strain state, with cylindrical implosion by explosives representing a triaxial state of deformation in contrast to the biaxial and uniaxial strain states moving to the right in the plot. Generally, martensite volume fraction increases with decreasing strain rate and increases with strain state at corresponding strain, for example, along a reference line at 20 % strain in Fig. 3b.

Since it is generally observed for many fcc and bcc metals and alloys that microindentation hardness (Vickers diamond) is simply related to the residual yield stress by

$$\sigma_y \cong H/3, \tag{6}$$

where H is the average measured hardness (Vickers hardness, HV, or VHN × 10^{-2} = GPa), deformation and hardness can also provide some sense of work hardening, and this is illustrated in Fig. 4 for plane (compression) shock-loaded metals and alloys. It can be noted in Fig. 4a that there is a significant increase in residual hardness over a range of metals and alloys all having comparable slopes except for

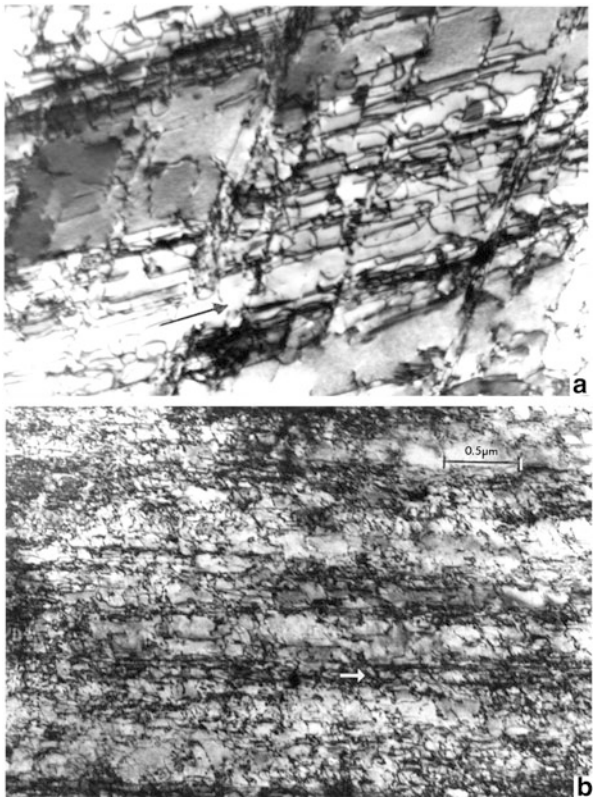


Fig. 2 Cold-rolled brass (70Cu–30Zn) observed by TEM. (**a**) 15 % reduction. (**b**) 30 % reduction. The surface orientation in (**a**) and (**b**) was (110). The *arrows* illustrate the $[1\bar{1}0]$ direction. Magnification for (**a**) and (**b**) is the same as given in (**b**)

ThO_2-dispersed Ni and ^{80}Ni–^{20}Cr (TD–Ni and TD–NiCr, respectively). In contrast, Fig. 4b illustrates work softening at high shock pressure for both 304 stainless steel and nickel where the energetically dense microstructures promote heating and subsequent recrystallization and related annealing and grain growth. It might be useful to refer back to the shock-induced microstructures and critical twinning pressures for their formation implicit in ▶ Figs. 7, ▶ 8, ▶ 9, ▶ 10, ▶ 11, ▶ 12, ▶ 13, and ▶ 14 of chapter "Planar Defects: Crystal Interfaces" and in contrast to ▶ Figs. 30, ▶ 31, and ▶ 32 of chapter "Line Defects: Dislocations in Crystalline Materials" as visual elements of the "collage building" alluded to previously. In this regard, it can be noted that not only is there a change from dislocation cells and stacking faults-to-deformation twins or twin faults for uniaxial, plane shock of fcc metals and alloys, but in some metals and alloys, the stress state (plane shock versus spherical shock (in crater formation alters the formation of twins in favor of microbands).

It might also be appreciated that many processes involving the forming of metals and alloys as well as their exposure to other extreme deformation conditions involve high strains and strain rates, as well as multiaxial strain states. For example, wire drawing often involves very high strains imposed at very high strain rates

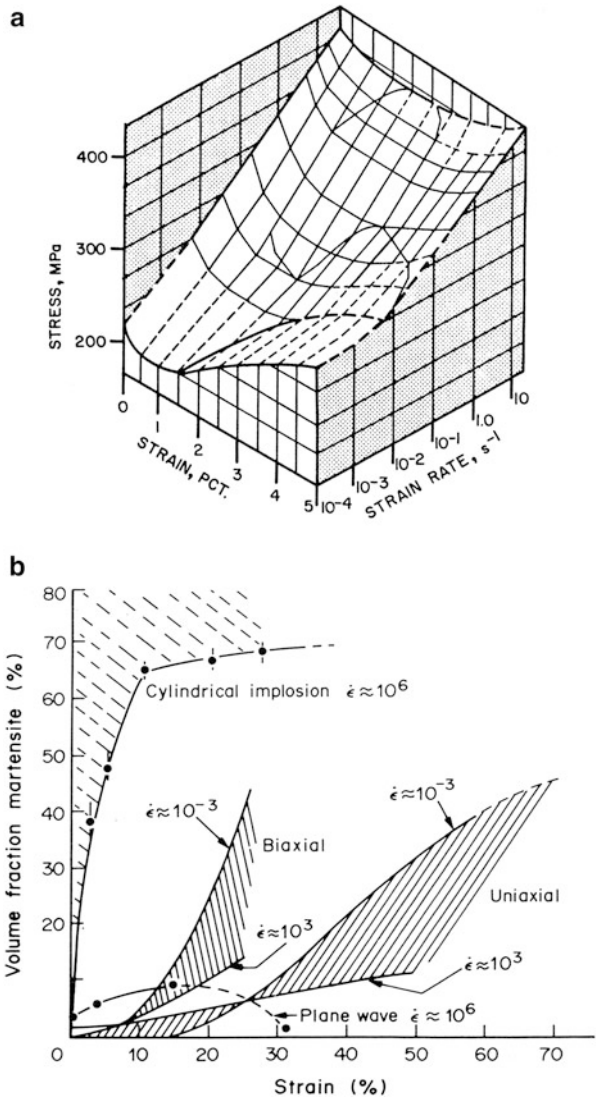


Fig. 3 (a) Stress–strain and strain rate plot for mild steel (adapted from Campbell (1972)). (**b**) α'-martensite volume fraction versus strain in 304 stainless steel for various strain rates and strain states. The shaded regions connect strain rate limits and characterize strain state regimes (Adapted from Johnson et al. (1985))

($>10^{-3}$/s) and in a nearly triaxial strain state. Correspondingly, high-speed forming and friction welding processes (particularly friction-stir welding) involve high strains and strain rates where the associated adiabatic heating (Eq. 3) changes the mode of deformation from dislocation emission, propagation (slip), and interaction to one involving dynamic recrystallization and superplastic flow of material in the solid state (by grain boundary sliding). This is also characteristic of ballistic or related projectile and target material interactions which are also characterized by dynamic recrystallization-induced solid-state flow. These phenomena will be treated more extensively in a later chapter.

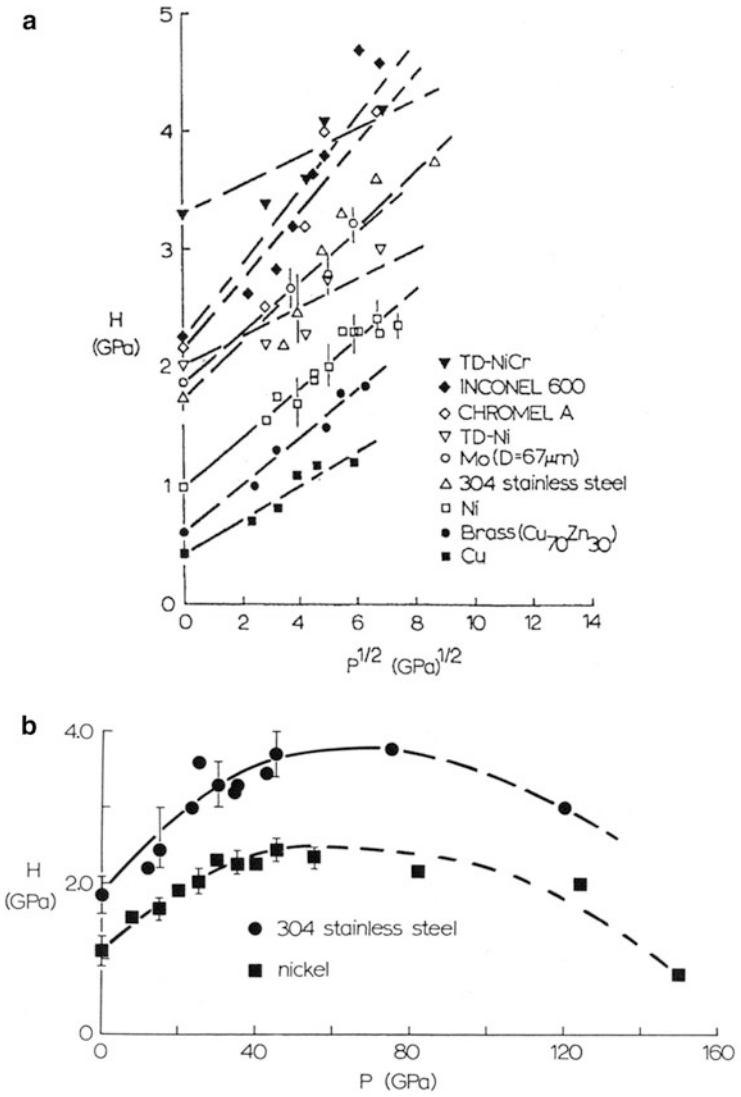


Fig. 4 (**a**) Hardness (Vickers) versus square root of plane shock pressure for metals and alloys. (**b**) Hardness versus shock pressure for Ni and 304 stainless steel over a wide pressure range

Imperfection Strengthening

For the most part, the work hardening or strain hardening described above has excluded grain size and the effect of grain or phase boundary volume fraction on ledge density and the dislocation density as represented in ▶ Eq. 11 of chapter "Line Defects: Dislocations in Crystalline Materials" Correspondingly, ▶ Eq. 13 of

chapter "Line Defects: Dislocations in Crystalline Materials" illustrates that the dislocation density is of course related to the grain size which, as indicated more explicitly in ▶ Eqs. 14 and ▶ 15 of chapter "Line Defects: Dislocations in Crystalline Materials" is related to the propensity of dislocation ledge sources in the grain boundary. Additionally, and as implicit in Figs. 1 and 2, the initial strength (represented by the residual hardness or yield stress) is often the result of prior deformation and the existence of dislocations or other defects, such as stacking faults (in low stacking-fault free energy metals and alloys) or twin faults.

One of the principal features of strengthening particularly in metals and alloys has involved so-called barriers to dislocation motion especially provided by planar defects and retarding dislocation motion by precipitates and other dispersed particles which can pin dislocations in various ways. The barriers to dislocation motion and material flow have historically invoked dislocation pileups, but it is certainly clear by examining the record provided herein (e.g., ▶ Figs. 18, ▶ 22, ▶ 27, ▶ 29 of chapter "Line Defects: Dislocations in Crystalline Materials") that pileups are as rare as Frank–Read sources and are virtually nonexistent. Effective work hardening involves the emission of dislocations especially at ledges in planar defects, interaction of dislocations and entanglement in the grain interior after emission, and their impedance by precipitates and dispersed phases.

Figure 5 illustrates a range of classical strengthening phenomena. For planar defects, beginning with grain boundaries partitioning a structure by grains of dimension D (Fig. 5b), strengthening is given by

$$\sigma = \sigma_0 + K\sqrt{D}, \tag{7}$$

where K is a constant usually dependent upon the shear modulus, G. But since the grain boundaries are dislocation sources, we could represent the strengthening or hardening simply in terms of ▶ Eq. 11 of chapter "Line Defects: Dislocations in Crystalline Materials":

$$\sigma = \sigma_0 + K\sqrt{\rho}, \tag{8}$$

or we might consider the consequence of adding the contributions of grain size and dislocation density in Fig. 5b, c as

$$\sigma = \sigma_0 + \left(K/\sqrt{D} + K'\sqrt{\rho}\right). \tag{9}$$

Studies of the contribution of order of magnitude change in grain size versus order of magnitude change in dislocation density, ρ, for ballistic impact of copper targets have shown that the actual contribution to hardening by the grain size, D, was roughly 10 % of hardness increase in contrast to ~90 % for the increase (by a factor of 10) in dislocation density (Murr et al. 1997). In this regard, many metals and alloys have been shown to be represented by an equation like Eq. 7, popularly referred to as the Hall–Petch equation (Hall 1951; Petch 1953). This is illustrated in Fig. 6. However, examining several specific deformation features in Fig. 6

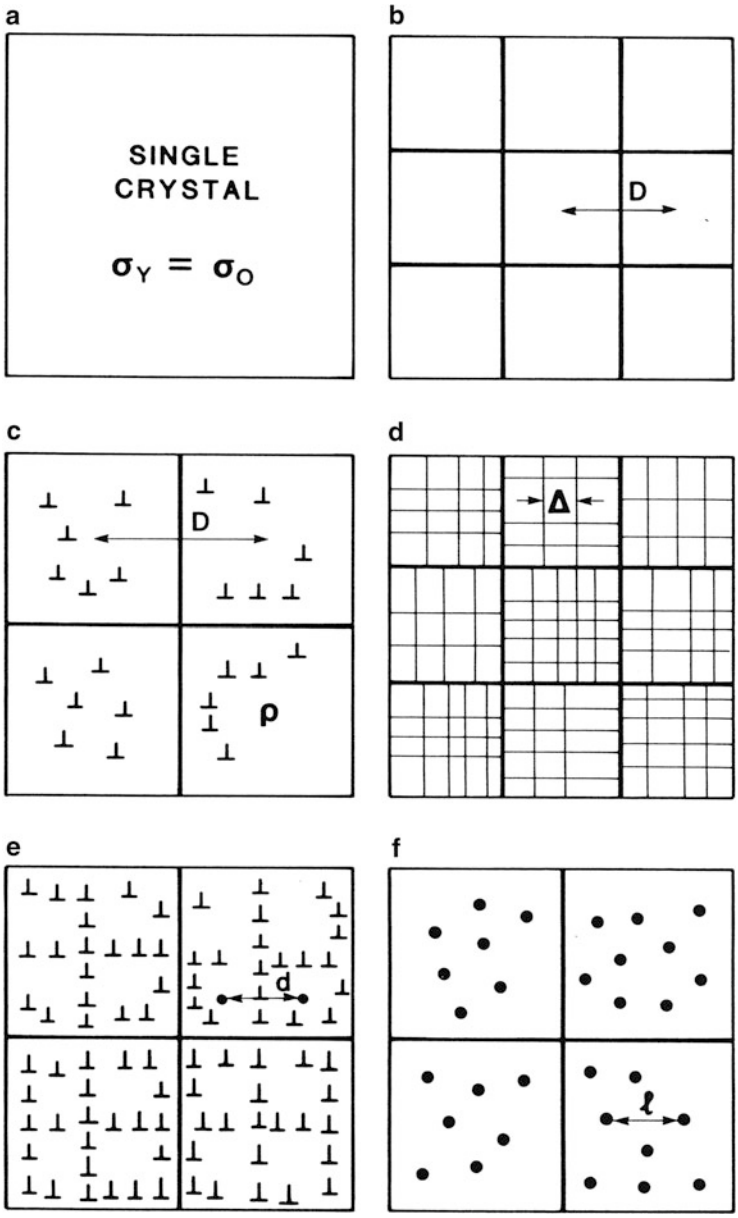


Fig. 5 Schematic views of microstructure or crystal imperfections contributing to strengthening. (**a**) Single crystal. (**b**) Polycrystalline material with grain size D. (**d**) Dislocations in polycrystalline materials with density ρ. (**d**) Twin/twin faults, stacking faults, or phase structures spaced Δ. (**e**) Dislocation cell structure spaced d. (**f**) Precipitates or dispersed phase spaced ℓ (or L). (Adapted from Murr 2000)

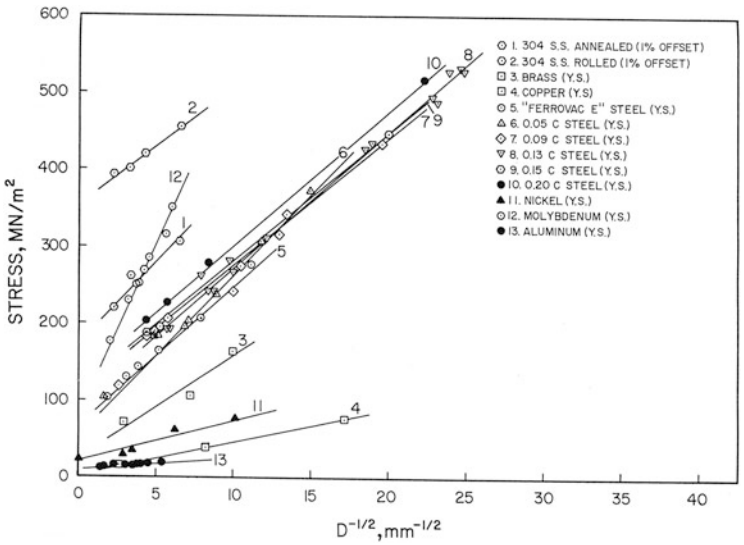


Fig. 6 Plot of residual stress (yield stress) versus reciprocal square root of grain size for metals and alloys (Adapted from Meyers and Chawla (1984); courtesy of K.C. Hsu)

reinforces the complexity of microstructural/imperfection contributions. For example, Fig. 6 shows two situations for 304 stainless steel indicated at 1 and 2, respectively. The annealed slope at (1) is identical to that for rolled stainless steel at (2), but the residual yield stress is increased (strengthened) by rolling (at some percent reduction). This incremental strengthening is due to dislocation production in the grains implicit in Fig. 6c, e as well as d, which would represent the production of martensite as illustrated in Fig. 3b. Consequently, we might represent the strengthening of the 304 stainless steel in Fig. 7 as:

$$\sigma = \sigma_0 + \left(K/\sqrt{D} + K'\sqrt{\rho} + K''/\sqrt{\Delta}\right), \tag{10}$$

where Δ represents the martensite spacing or even volume fraction in some related form. The strengthening increment in Fig. 6 for stainless steel is then

$$\left(K'\sqrt{\rho} + K''/\sqrt{\Delta}\right), \tag{11}$$

and this increment beyond just the grain size, D, accounts for an increase in residual stress of roughly 88 %. This is consistent with the variance in grain size strengthening versus dislocation strengthening noted above for impact deformed Cu (Murr et al. 1997), although this increment includes dislocation density and martensite volume fraction. It might also be recalled that the dislocation density and dislocation cell size, d, are related as illustrated in ► Fig. 32b of chapter "Line Defects: Dislocations in Crystalline Materials" Consequently, where dislocation

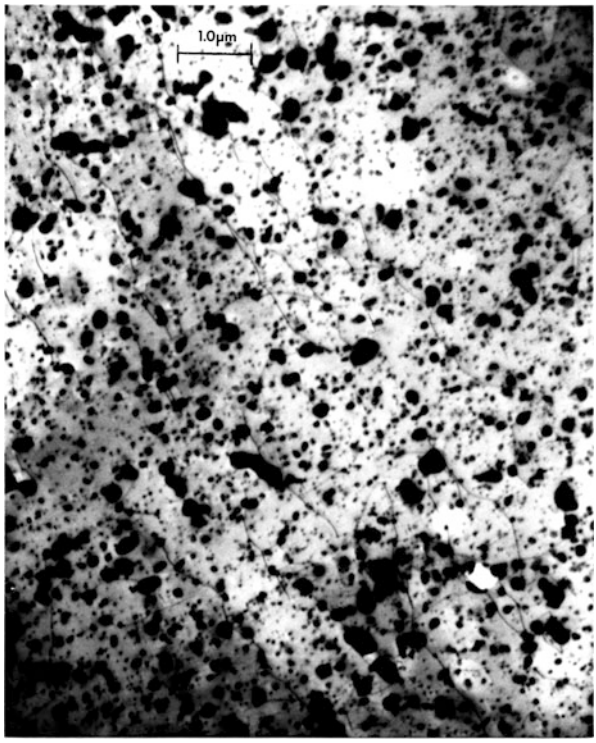


Fig. 7 TEM image of dispersion-hardened nickel (2 vol.% ThO_2). Note long dislocation lines appearing as threads, pinned at particles

cells form in high stacking-fault free energy metals, $\sqrt{\rho}$ can be replaced by 1/d in ▶ Eq. 15 of chapter "Line Defects: Dislocations in Crystalline Materials" or Eq. 9.

It might also be recognized that for Δ in Eq. 10 representing stocking faults, the functionality will probably revert to $1/\Delta$ rather than $1/\sqrt{\Delta}$, and twin faults may in fact vary between these functions. This may also be the case for microbands as well.

On examining the strengthening by dislocation pinning by precipitates or dispersions such as ThO_2 single-crystal particles in Ni or NiCu (^{80}Ni–^{20}Cu), the incremental strengthening can be observed in Fig. 4a for plane shock loading. For example, the hardening observed for shocked Ni follows a slope identical to that for Td–Ni (2 % by volume ThO_2 added to Ni), just as the case for the grain size slope in Fig. 6 for 304 stainless steel. For the dispersion-strengthen Ni in Fig. 4a, we can write

$$\sigma = \sigma_o + \left(K/\sqrt{D} + K'\sqrt{\rho} + K''/L\right), \tag{12}$$

where L in Eq. 12 represents the interparticle spacing shown in Fig. 5f. In the comparative hardness increments (slope) for Ni and TD–Ni in Fig. 4a, we might consider the increment only for the dispersed phase, K″/L, which accounts for a hardness change of 1–2 GPa, or incremental hardness change of 100 %, a significant strengthening increment.

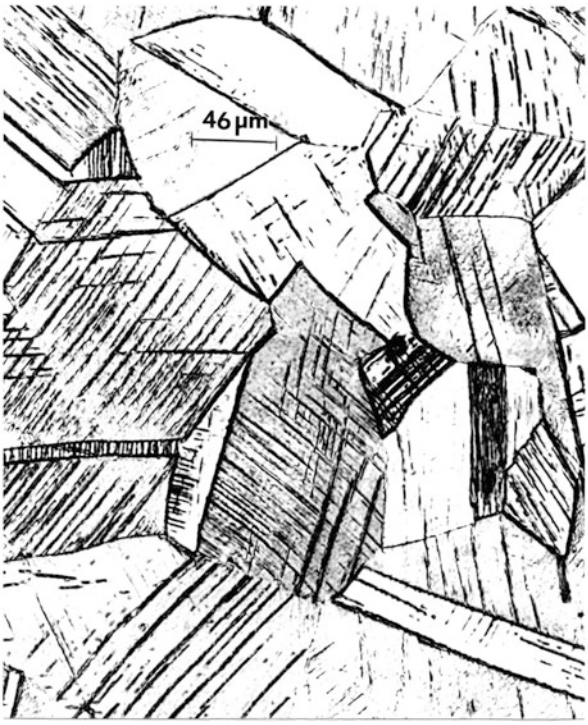


Fig. 8 Twin faults in large grains of oblique shock-loaded Cu

Figure 7 shows the basic TD–Ni microstructure with 2 % (volume) ThO_2. Dislocations pinned to ThO_2 particles are even evident in this image, providing some indication of the strengthening efficiency of dispersion hardening. However, the efficiency of strengthening declines with a decline in volume fraction since L (Fig. 5f) is increased (Eq. 12), and at some volume fraction greater than ~2 % ThO_2, the efficiency would also decline even though ideally L would decrease. In addition, the dispersion efficiency may be different in different matrices, which is apparent on comparing the incremental hardness change for chromel (NiCr) versus TD–NiCr, which has the same ThO_2 dispersion as shown in Fig. 7.

It might also be noted that grain size can be limiting in terms of stacking fault or twin fault production because of dislocation flow limitations. Consequently, large grain sizes will produce a larger volume fraction of stacking faults or deformation-induced twin faults as illustrated in Fig. 8 for Cu. Figure 9 illustrates this quantitatively for Cu for which Fig. 8 illustrates the largest grain size and associated deformation (strain) in Fig. 9 [186]. In the context of defect strengthening, it can be observed that in the situation represented by Fig. 9, increasing grain size is offset by increasing twin (or twin fault) volume fraction:

$$\left(K/\sqrt{D} + K'/\Delta\right) \tag{13}$$

in Eq. 10.

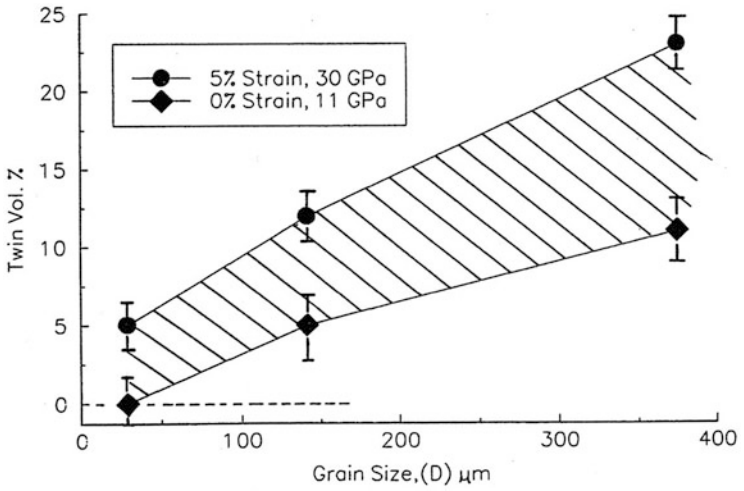


Fig. 9 Twin (twin fault) volume fraction (%) versus grain size for oblique shock-loaded Cu. Shading emphasizes shock-related variations (Adapted from Sanchez et al. (1997))

As illustrated in ▶ Figs. 3 and ▶ 4 of chapter "Volume Defects: 3D Imperfections in Crystals" the systematic development of precipitation in various forms by aging or related heat treatment constitutes another category of strengthening by imperfections. This can involve pinning center as represented by L in Fig. 5f and Eq. 12 or in the case of precipitate plates or Widmanstätten structures which are spaced Δ as in Eq. 10 (Fig. 5d) and illustrated in ▶ Fig. 4b, in chapter "Volume Defects: 3D Imperfections in Crystals"

It can be appreciated, in retrospect, that strengthening of materials is a complex yet strategic feature of materials science and engineering as it applies to materials processing and performance expectations. Predictive algorithms are difficult to devise because, as illustrated in Eqs. 7–12, there is a complex interplay between defects and their contribution to strengthening and hardening, and there is far too little data to provide a comprehensive understanding that would permit broad predictions, especially with variations in strain, strain rate, and temperature implicit in broad data constructs shown in Fig. 3a. Nonetheless, we have hopefully provided in this overview a compelling collage of strengthening fundamentals which can provide some meaningful guidance in understanding materials strengthening and hardening as these apply to expected and observed performance issues.

It is perhaps obvious, in retrospect, to realize that crystal imperfections are the essence of materials structure and microstructure and perhaps the most important aspects of materials properties and performance. Only voids pose any direct adversities regarding a material's integrity, while, of course, other imperfections can act as precursors for cracks or other forms of degradation such as corrosion,

stress corrosion cracking, or other compromises to materials behavior. It has already been illustrated that defects determine color and other important properties of crystals as well as their associated optical and electronic properties. In this chapter, it has been observed how defects control mechanical response, strength, hardness, and related properties and performance issues. Although in the more general sense, imperfections may be viewed as an adversity, in materials, they are the essence of design in relationship to properties and performance. In the four chapters in Part VI, following properties of materials in relation to imperfections and especially the measurement and prediction of properties in single crystals will be examined.

The concept of disclinations as they may either substitute for dislocations or provide a more tractable representation of the defect or defect microstructure was discussed in previous chapters. This is especially true for quasicrystals discussed in chapter "► Crystallography Principles," where normal crystallographic slip planes and systems are not represented, and wedge phenomena are the norm in dealing with fivefold symmetry.

It is important to note that dislocation production is not represented by Frank–Read sources and these are almost never observed in deformed materials. Furthermore, the concept of dislocation pileups is also largely a misrepresentation of dislocation emission, and the absence of grain matrix sources of dislocations negates the necessity for, or the reality of, dislocation pileups. Consequently, the more practical if not pragmatic view of strengthening in materials has more to do with ledge sources at interfaces and the resulting dislocation reactions and interactions which provide strength by limiting glide of dislocations as discussed in chapter "► Line Defects: Dislocations in Crystalline Materials"

Although precipitation- and dispersion-hardened and precipitation- and dispersion-strengthened materials have been discussed, this discussion has not focused specifically on composite materials, metal matrix composites (MMCs), or fiber-reinforced materials, which are an important aspect of materials strengthening. Composites are discussed in Part VIII devoted to aspects of those materials and materials systems.

References

Campbell JD (1972) Dynamic plasticity of metals. Springer, Vienna

Hall EO (1951) The deformation and aging of mild steel : III discussion and results. Proc Roy Soc B64:474

Johnson KA, Murr LE, Staudhammer KP (1985) Comparison of residual microstructures for 304 stainless steel shock loaded in plane and cylindrical geometries: implications for dynamic compaction and forming. Acta Metall 33(4):677–684

Meyers MA, Chawla KK (1984) Mechanical metallurgy : principles and applications. Prentice-Hall, Englewood Cliffs

Murr LE (2000) Industrial materials science and engineering, Chap. 6.6. In: Shell RL, Hall EL (eds) Handbook and industrial automation. Marcel Dekker, New York, pp 529–566

Murr LE (1975) Interfacial phenomena in metals and alloys. Addison-Wesley Publishing Co., Reading. Crater-related microstructures in thick copper targets: part 2. J Mater Sci 32: 3147–3156

Murr LE, Niou CS, Ferreyra E, Garcia EP, Liu G, Horz F, Bernhard RP (1997) Effect of initial microstructure on high velocity and hypervelocity impact cratering and crater-related microstructures in thick copper targets :part 2. J Mater Sci 32:3147–3156

Petch NJ (1953) The cleavage strength of polycrystals. J Iron Steel Inst 174: 25–28

Sanchez J, Murr LE, Staudhammer KP (1997) Effect of grain size and pressure on twinning and microbanding in oblique shock loading of copper rods. Acta Mater 45(8):3223–3235

Part VI

Crystal Physics: Tensor Representation for Physical Properties of Crystalline Materials

Examples of Tensor Properties Using Matrix Fundamentals (A Physical Property)

Contents

Abstract

Properties of single crystals in particular often differ with different crystal directions as a consequence of constraints imposed by atomic packing and arrangements as well as the placement of substitutional impurities or other defects. For polycrystalline materials with more random orientations of crystal grains, these properties usually average out, and measurements represent the same values in any direction. Such multi-directional or multi-vector properties are intrinsically connected, and these connected properties are represented by tensors or arrays of vectors represented crystallographically (as crystal directions). These tensors or vector arrays have a rank characteristic of their complexity, which manifests itself in the scalar coefficients characteristic of a property. Such notations are conveniently represented by matrix algebra which in the case of tensor properties represents a powerful mathematical tool to study or predict properties of crystals, particularly crystal orientations.

Introduction

The concept of a property (a physical property) involves a measure of the response of structure to some imposed condition or action (such as temperature, pressure, stress, etc.) or the alteration of, or interactions of, waves or particles with the structure. In the ideal case of crystal structure, the condition or action is imposed on the lattice and interaction of waves or "particles" – atoms in the lattice, electrons, photons, and phonons – interacting with lattice atoms: their atomic number, size,

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_19

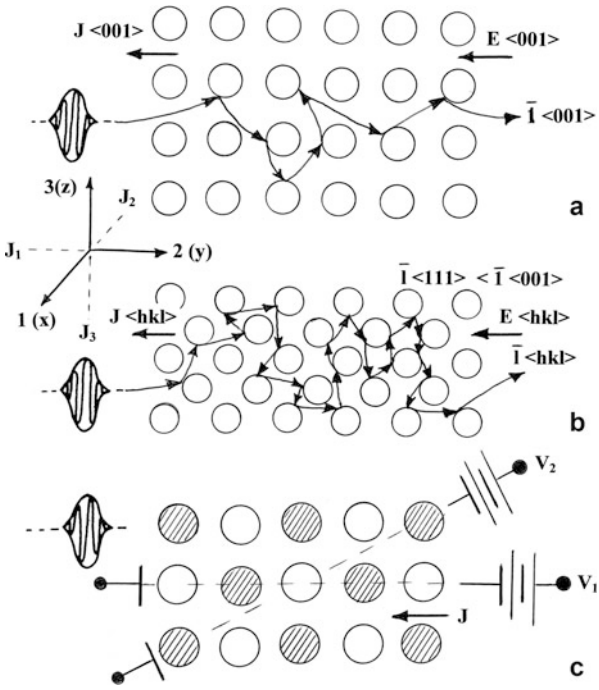


Fig. 1 Schematic representation of mean free path variation with lattice structure and/or direction. (**a**) Simple cubic symmetry. (**b**) Symmetry and direction variation $< hkl >$. (**c**) Binary structure with voltage and electric field (E_i) parallel and nonparallel to J

arrangement (or symmetry), bonding, and concentration in multi-atom or multi-element structures. As a consequence, the measurement of a physical property is likely to vary in different crystal directions or structures and to be influenced by atom specie and size and the degree of crystallinity or the size and orientation of a crystal domain, as well as imperfections.

These features, particularly the effect of lattice symmetry or orientation on a physical property such as conductivity or resistivity, are shown schematically (and intuitively) in Fig. 1. In this example (Fig. 1), the effect of lattice atom arrangement and atomic species (or size) and the application of an applied electric field, **E**, on the electron mean free path, ℓ, (or I) is apparent. For an orientation presenting a more open arrangement where a longer mean free path is favored, the conductivity is given by:

$$\sigma = \frac{n_c e^2}{m\bar{v}}\left(\bar{\ell}\right) \tag{1}$$

or the corresponding resistivity, ρ, is given by ρ (resistivity) $= 1/\sigma$:

$$\rho = \frac{m\bar{v}}{n_c e^2}\left(\frac{1}{\bar{\ell}}\right), \tag{2}$$

This is observed in Fig. 1a which when viewed in a crystal might represent the resistivity in the <001> direction: $\overline{\ell} = \overline{\ell} < 001>$. In contrast, the mean free path in a close-packed direction as represented by Fig. 1b will be shorter because of the more frequent interaction of the electron with the lattice atoms, $\overline{\ell} < hk\ell > < \overline{\ell} < 001>$, and the resistivity will increase for the condition implied by Fig. 1b. Correspondingly, Fig. 1c shows that even for a specific orientation, or a multi-component and ordered structure, the application of an electric field, **E**, to move the electron (or electrons, n_c in Eq. 2), creates a current or current density, **J**, having scalar values in current (amperes)/unit area (cm^2). But **E** and **J**($\mathbf{E} = \rho\mathbf{J}$) are not necessarily parallel, in which case ρ is not a scalar value. This situation is represented in Fig. 1c where the electric charge (current) flows more easily in a direction which differs from that in which the electric field is applied, corresponding to differences in resistivity as a consequence of differences in the mean free path, $\overline{\ell} < hkl >$, with crystal directions <*hkl*>. Changes in the atomic species (and possibly size) also change the mean free path.

As illustrated in Fig. 2, the electric field can be resolved into components of the vector **E** in Cartesian coordinates, where we also assume linearity between the electric field components and the current density components. This can be expressed mathematically by:

$$\left.\begin{aligned} E_1 &= \rho_{11}J_1 + \rho_{12}J_2 + \rho_{13}J_3 \\ E_2 &= \rho_{21}J_1 + \rho_{22}J_2 + \rho_{23}J_3 \\ E_3 &= \rho_{31}J_1 + \rho_{32}J_2 + \rho_{33}J_3 \end{aligned}\right\} \tag{3}$$

Equation 3 can be alternatively represented in matrix (algebra) form:

$$\begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} = \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{12} & \rho_{13} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{pmatrix} \begin{pmatrix} J_1 \\ J_2 \\ J_3 \end{pmatrix} \tag{4}$$

Equation 4 can also be expressed as

$$E_i = \sum_{i=1}^{3} \rho_{ij}J_j = \rho_{ij}J_j \tag{5}$$

or

$$\mathbf{E} = \boldsymbol{\rho}\mathbf{J} \tag{6}$$

In Eq. 4, the matrices

$$\begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} \text{ and } \begin{pmatrix} J_1 \\ J_2 \\ J_3 \end{pmatrix}$$

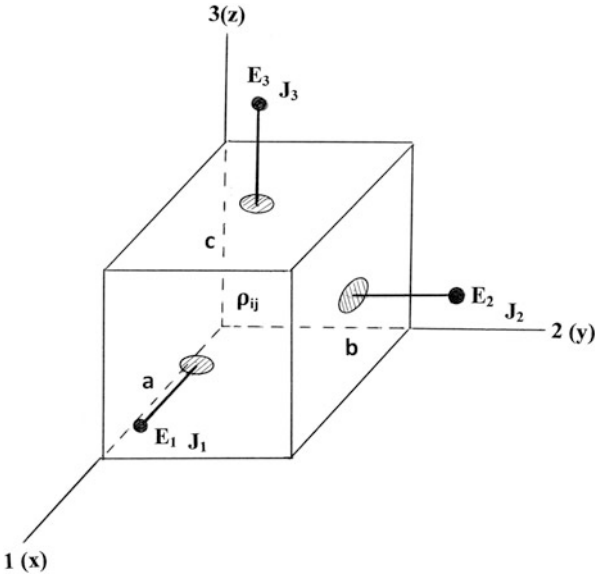


Fig. 2 Vector components, E_i, J_i applied to an orthogonal crystal unit cell ($\alpha = \beta = \gamma = 90°$) where Ohm's law applies within the unit cell: $E_i = \rho_j J_j$

are column vectors, while

$$\begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{12} & \rho_{13} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{pmatrix}$$

is a 3 × 3 or square matrix of coefficients and is a so-called second-rank tensor, where the ρ_{ij} terms are coefficients of the tensor and have scalar values. Correspondingly, values of E_i and J_j are coefficients or scalar values of the corresponding vectors or tensors of rank one. The index notation denotes the tensor rank, I or j = > rank 1 (a vector), ij = > rank 2, ijk = > rank 3, ijkl = > rank 4, etc.

Tensors can be described as multi-linear, scalar-valued functions of vectors, as illustrated in Eqs. 4 and 5. A tensor is an operator and stores information in the form of a linear mapping of vectors. In this sense, tensors feed on vectors to produce numbers which are scalar coefficients describing a material property. It must be pointed out that for symmetrical tensors, the tensor coefficients of a property are related to a crystal structure unit cell implicit in Fig. 2. For example, the diagonal coefficients for resistivity in Eq. 4 represent the principal values in corresponding directions, $\rho_{11} = > \rho_x$; $\rho_{22} = > \rho_y$; $\rho_{33} = > \rho_z$, where x, y, z (or 1, 2, 3) imply crystal directions [100], [010], and [001], respectively. Figure 2 therefore represents the crystal basis for properties, and where property coefficients are required in a specific orientation, [hkl], perhaps for growth of a crystal in an orientation [hkl], this basis must be properly oriented and the tensor matrix appropriately transformed. Such transformations are accomplished by a transformation matrix, which will be described and illustrated later in this chapter.

It might also be pointed at that while isotropic cubic crystals have equivalent properties in all directions when the equivalent tensor is second rank, higher-rank tensors do not exhibit this behavior. This is particularly true of elastic properties, stiffness, and compliance coefficients, to be described later in this chapter, where it will be shown that the number of these coefficients for cubic single crystals is different in contrast to polycrystalline materials having random grain orientations.

The classic work of J.F. Nye (1959) describes physical properties of crystals using tensors and can serve as a comprehensive overview of physical properties of crystalline material. In this regard it might be noted that anisotropic properties of crystalline materials cannot be measured or represented as tensors. These include tensile strength, flow stress, crystal and grain growth, interfacial free energies, etc. Properties such as density, temperature, etc. are represented by zero-rank tensors which are characteristically scalar values.

The use of tensor formalism to investigate, calculate, or even predict the properties of materials is a powerful as well as convenient tool set in the materials sciences and engineering. In this presentation we will address only a few concepts to allow the reader some appreciation of this approach.

In general, a matrix is any rectangular array of numbers (or coefficients having scalar value):

$$\mathbf{T} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \tag{7}$$

where **T** is a matrix (or tensor) with m rows and n columns, having an order m $\times$ n. An m $\times$ 1 matrix is called a column matrix as illustrated above in Eq. 4, whereas a 1 $\times$ n matrix is called a row matrix which is represented in Eq. 4 as ($\rho_{11}\rho_{12}\rho_{13}$) *or* [$\rho_{11}\rho_{12}\rho_{13}$]. In this sense it should be noted that brackets or parens are identical notations and will be used interchangeably in this chapter.

It might be suspected that for a regular crystal structure – a single crystal structure – we would not have nine independent components of resistivity represented in Eq. 4. Indeed, because of symmetries and variations in lattice parameters and direction cosines, the relationship between second-rank tensors and crystal systems (▶ Fig. 2 in chapter "Crystallography Principles") becomes

$$\text{Cubic} \rightarrow \begin{bmatrix} T & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{bmatrix}$$

$$\begin{matrix} \text{Tetragonal} \\ \text{Hexagonal} \\ \text{Trigonal} \end{matrix} \rightarrow \begin{bmatrix} T_1 & 0 & 0 \\ 0 & T_1 & 0 \\ 0 & 0 & T_3 \end{bmatrix}$$

$$\text{Orthorhombic} \rightarrow \begin{bmatrix} T_1 & 0 & 0 \\ 0 & T_2 & 0 \\ 0 & 0 & T_3 \end{bmatrix}$$

$$\text{Monoclinic} \rightarrow \begin{bmatrix} T_{11} & 0 & T_{13} \\ 0 & T_{22} & 0 \\ T_{13} & 0 & T_{33} \end{bmatrix}$$

$$\text{Triclinic} \rightarrow \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix}$$

where in the context of the resistivity tensor in Eq. 4, $\rho_{ij} = T_{ij}$ and ρ_{11}, ρ_{22}, $\rho_{33} \equiv T_{11}$, T_{22}, $T_{33} \equiv T_1$, T_2, T_3, etc. We observe for cubic symmetry that all of the diagonal coefficients are equal, $T_1 = T_2 = T_3 = T$, representing a perfectly isotropic crystal structure. Consequently, the resistivity tensor in Eq. 4 for cubic crystals can be written

$$\begin{bmatrix} T & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{bmatrix} = \begin{bmatrix} \rho & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & \rho \end{bmatrix}$$

and the resistivity is the same in all directions. On the other hand, this assumes a single-component structure. For an ordered versus a disordered binary system as illustrated in ▶ Fig. 1 of chapter "Structure of Metals and Alloys," the resistivity and conductivity will change. In addition, for an ordered AB structure in contrast to A_3B, the tensor could be represented by $T_1 = T_2 \neq T_3$ as for a tetragonal crystal structure.

Of course the conditions for tensor (matrix) coefficients of a property in crystals as expressed above are generally applicable for any property described by a second-rank tensor: electrical and thermal resistivity and conductivity, permittivity, permeability, and diffusion which and all second-rank symmetrical tensors represented by Eqs. 4 and 5. In addition, properties which are inverse, such as resistivity and conductivity, require a specific inverse operation for matrices:

$$T \cdot T^{-1} = T^{-1} \cdot T = [1], \tag{8}$$

where [1] is referred to as the unity matrix:

$$\begin{aligned} [\delta_{ij}] &= [1], \\ \delta_{ij} &= 1; i = j; \delta_{ij} = 0; i \neq j \end{aligned} \tag{9}$$

where δ_{ij} is referred to as the *Kronecker Delta*. For a second-rank tensor such as resistivity and conductivity, Eq. 8 is written as

$$\begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{bmatrix} \begin{bmatrix} \rho_{11}^{-1} & \rho_{12}^{-1} & \rho_{31}^{-1} \\ \rho_{21}^{-1} & \rho_{22}^{-1} & \rho_{32}^{-1} \\ \rho_{31}^{-1} & \rho_{32}^{-1} & \rho_{33}^{-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{10}$$

or since

$$\begin{bmatrix} \rho_{11}^{-1} & \rho_{12}^{-1} & \rho_{31}^{-1} \\ \rho_{21}^{-1} & \rho_{22}^{-1} & \rho_{32}^{-1} \\ \rho_{31}^{-1} & \rho_{32}^{-1} & \rho_{33}^{-1} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \rho_{33} \end{bmatrix}, \tag{11}$$

where σ_{ij} is the conductivity tensor, Eq. 10 can be expressed by

$$\begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

Equation 12, when multiplied and evaluated, becomes:

$$\left.\begin{aligned} \rho_{11}\sigma_{11} + \rho_{12}\sigma_{21} + \rho_{13}\sigma_{31} &= 1 \\ \rho_{11}\sigma_{12} + \rho_{12}\sigma_{22} + \rho_{13}\sigma_{32} &= 0 \\ \rho_{11}\sigma_{13} + \rho_{12}\sigma_{23} + \rho_{13}\sigma_{33} &= 0 \\ \rho_{21}\sigma_{11} + \rho_{22}\sigma_{21} + \rho_{23}\sigma_{31} &= 0 \\ \rho_{21}\sigma_{12} + \rho_{22}\sigma_{22} + \rho_{23}\sigma_{32} &= 1 \\ \rho_{21}\sigma_{13} + \rho_{22}\sigma_{23} + \rho_{23}\sigma_{33} &= 0 \\ \rho_{31}\sigma_{11} + \rho_{32}\sigma_{21} + \rho_{33}\sigma_{31} &= 0 \\ \rho_{31}\sigma_{12} + \rho_{32}\sigma_{22} + \rho_{33}\sigma_{32} &= 0 \\ \rho_{31}\sigma_{13} + \rho_{32}\sigma_{23} + \rho_{33}\sigma_{33} &= 1 \end{aligned}\right\} \tag{13}$$

This results of course in nine equations and nine unknowns, which can be solved for either ρ_{ij} or σ_{ij}, respectively. Note that the multiplication of the matrices is row-column multiplication of the matrix (tensor) coefficients, which equal the corresponding unity matrix (or tensor) coefficient, 0 or 1. Note also that the standard units of the resistivity coefficients, ρ_{ij}, are in ohm-cm (Ω-cm), while conductivity coefficients have units of 1/Ω-cm, often called mhos (reciprocal ohm-cm). As illustrated above, Eq. 13 only applies to triclinic crystals, while for ideal (single-component) cubic crystals, ρ and σ are the same everywhere.

Diffusion as presented earlier in ▶ Eq. 4 (chapter "Point Defects") is represented by a second-rank tensor as the symmetrical diffusivity matrix:

$$\sigma_i = D_{ij} \frac{\partial C}{\partial x_i}, \tag{14}$$

where D_{ij} corresponds to a specific temperature for diffusion of a specie with concentration, C. In single crystals, the diffusion specie may not diffuse parallel

Table 1 Thermal expansion/contraction components for a few crystal structures

	Thermal expansion/contraction coefficients		
Crystal	α_1	α_2	α_3
Al (fcc)	22	22	22
Fe (bcc)	12	12	12
Calcite ($CaCo_3$) (trigonal)	−5.6	−5.6	25
Gyprum ($Ca\delta O_4$) (monoclinic)	1.6	42	29
Gypson ($CaSo_4$) (random)	24	24	24
Si_3N_4 (hexagonal)	1.3	1.3	2.8

to the concentration gradient, $\partial C/\partial x$, as implicit in the representation of Eq. 13, the tensor form of Fick's law.

Similarly, and for an isotropic material, the dielectric properties can be expressed by a second-rank tensor:

$$\mathbf{D} = \epsilon_0 \mathbf{KE} = \boldsymbol{\epsilon} \mathbf{E}, \tag{15}$$

where ϵ_0 is the permittivity of free space, $\mathbf{K}$ is the dielectric constant tensor, and $\boldsymbol{\epsilon}$ is the permittivity tensor. However, for a non-isotropic material,

$$D_i = \epsilon_o K_{ij} E_j = \epsilon_{ij} E_j, \tag{16}$$

where K_{ij} and ϵ_{ij} are symmetrical second-rank tensors having the form of Eq. 4.

It might also be noted that properties such as thermal expansion, α, involve a second-rank tensor relating a scalar and another second-rank tensor. In this case, the thermal expansion coefficient relates the contraction or expansion (strain, ϵ) to change in temperature ΔT:

$$\epsilon_{ij} = \alpha_{ij} \Delta T, \tag{17}$$

where ϵ_{ij} and $\propto_{ij}$ are symmetrical tensors as illustrated in Eq. 4. Table 1 illustrates a few examples for diagonal matrix components of $\alpha_{ij} = \alpha_i$ (i = 1, 2, 3) which are consistent with the crystal tensors noted previously. Other properties of crystalline materials are also represented by third-and fourth-rank tensors: T_{ijk}, T_{ihkl}. Third-rank tensors relate a vector to a second-rank tensor as illustrated, for example, by the piezoelectric effect where a stress (or pressure), σ, is applied to an asymmetrical crystal to create an electric dipole moment, P, which is proportional to the stress:

$$P = \Delta \sigma, \tag{18}$$

where Δ is the piezoelectric modulus which is positive for a tensile stress and negative for a compressive stress:

$$P_i = \Delta_{ijk} \sigma_{jk}. \tag{19}$$

The converse of the piezoelectric effect occurs when the application of an electric field, E, produces a displacement or strain of the crystal:

$$\epsilon_{jk} = \Delta_{ijk} E_i, \tag{20}$$

where E_i is a vector as in Eq. 5. Equation 20 couples electric and elastic behavior of piezoelectric crystals.

In the general case, a third-rank tensor will have 27 components or coefficients, but both σ_{jk} and ϵ_{jk} in Eqs. 19 and 20 above are symmetric in jk. Consequently, Δ_{ijk} will also be symmetric in jk, which reduces the coefficients to 18 and represented by a 6×3 matrix. Consequently Eqs. 19 and 20 can be written in the following subscript notation:

$$\left.\begin{aligned} P_m &= \Delta_{mn}\sigma_n \\ \epsilon_m &= \Delta_{mn} E_n \end{aligned}\right\} \tag{21}$$

where $m = 1, 2, 3$ and $n = 1, 2 \ldots 6$, and

$$\begin{aligned} \Delta_{ijk} &= \Delta_{mn} \quad \text{for } n = 1, 2, 3 \\ 2\Delta_{ijk} &= \Delta_{mn} \quad \text{for } n = 4, 5, 6, \end{aligned}$$

and the piezoelectric modulus matrix will take the form

$$\begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{bmatrix},$$

where d_{mn} does not transform as a tensor, whereas d_{ijk} does. Tensor transformation will be discussed later.

Elastic Properties of Crystals

In the case of fourth-rank tensors, two second-rank tensors are related. The most important example is stress, $\boldsymbol{\sigma}$, and strain, $\boldsymbol{\epsilon}$:

$$\sigma_{ij} = C_{ijkl}\epsilon_{kl} \tag{22}$$

or

$$\epsilon_{ij} = S_{ijkl}\sigma_{kl}, \tag{22}$$

where Eq. 22 represents the tensor form of Hooke's law,

$$\sigma = E\epsilon, \tag{24}$$

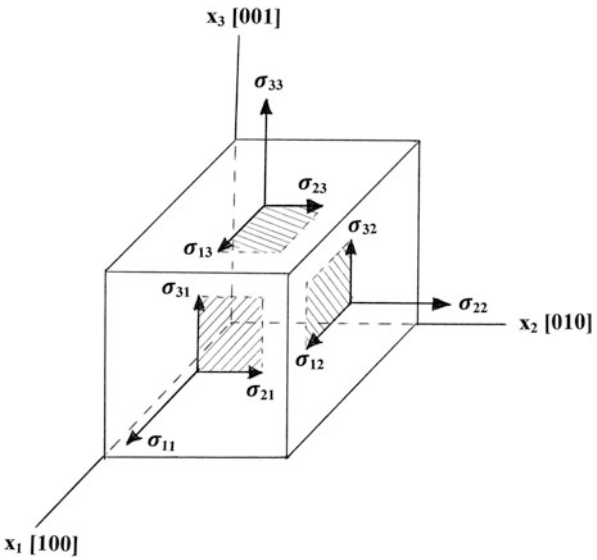


Fig. 3 Representation of the stress (and strain) tensor coefficients for cubic crystals (unit cells)

and E is the Young's modulus (or the elastic modulus). In the tensor form of Eq. 22, C_{ijkl} represents a 9 × 9 matrix with 81 coefficients which are called the *elastic stiffness coefficients* or *elastic constants*. Correspondingly, S_{ijkl} represents the inverse tensor also composed of 81 *elastic compliance coefficients*.

Figure 3 illustrates the application of normal stress (or strain) either in tension (extension) or compression and the related shear stresses resolved in the faces of a cubic crystal unit cell. The resulting deformation of will depend upon the actual direction of stress application and the ease of slip in specific directions: <110 > for fcc and <111 > for bcc as noted in the chapter "▶ A Brief Introduction to Quantum Mechanics." The corresponding stress tensor, σ_{ij}, can be written as

$$\sigma_{ij} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{212} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}, \tag{25}$$

which is equivalent to the tensor form of the strain, ϵ_{ij}. From the generalized form of Hooke's law in Eq. 24, we can write

$$\epsilon_{11} = \sigma_{11}/E; \epsilon_{22} = \sigma_{22}/E; \epsilon_{33} = \sigma_{33}/E, \tag{26}$$

and

$$\epsilon_{11} = \epsilon_{22} = -\upsilon\sigma_{33}/E, \tag{27}$$

where υ is Poisson's ratio

$$\upsilon = -\epsilon_{11}/\epsilon_{33} = -\epsilon_{22}/\epsilon_{33}. \tag{28}$$

The matrix C_{ijkl} in Eq. 22, relating to the nine coefficients of stress, (σ_{ij}), shown in Eq. 25 to the nine coefficients of strain, ϵ_{ij}, is symmetrical about the 9×9 matrix diagonal:

$$C_{ijkl} = C_{jikl}$$

and

$$C_{ijkl} = c_{ijlk}.$$

These elastic coefficients can be written in a contracted matrix notation, C_{mn}, similar to the piezoelectric modulus in Eq. 21, where in this case m and n are each indices corresponding to the pair of indices ij or kl, according to the following:

Tensor notation : ij or kl	11	22	33	23	31	12	32	31	21
Matrix notation : m or n	1	2	3	4	5	6	7	8	9

(29)

Consequently, by definition we can write

$$C_{11} = C_{1111}; C_{12} = C_{1122} = C_{2211} = C_{21}$$
$$C_{44} = C_{2323}; C_{46} = C_{2312}$$

and because of symmetry in cubic crystals, Eq. 22 can be written as

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{31} \\ \sigma_{12} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{12} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{13} & C_{23} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{14} & C_{24} & C_{34} & C_{44} & C_{45} & C_{46} \\ C_{15} & C_{25} & C_{35} & C_{45} & C_{55} & C_{56} \\ C_{16} & C_{26} & C_{36} & C_{46} & C_{56} & C_{66} \end{bmatrix} \begin{bmatrix} \epsilon_{11} \\ \epsilon_{22} \\ \epsilon_{33} \\ \epsilon_{23} \\ \epsilon_{31} \\ \epsilon_{12} \end{bmatrix} \tag{30}$$

where in Eq. 25, the shear stress components are $\sigma_{12} = \sigma_{21}$, $\sigma_{31} = \sigma_{13}$, $\sigma_{23} = \sigma_{32}$ (Fig. 3). The corresponding shear strain components can also be written as $\epsilon_{12} = \epsilon_{21} = \gamma_{12}$, $\epsilon_{31} = \epsilon_{13} = \gamma_{31}$, $\epsilon_{23} = \epsilon_{32} = \gamma_{23}$, where the γ_{mn} are the shear strains. The C_{mn} in the 6×6 matrix in Eq. 30 can be reduced to 21 independent elastic constants because of crystal symmetry, and as reflected in ▶ Eq. 5 (chapter "Planar Defects: Crystal Interfaces"), because the coordinate axes in Fig. 3 coincide with the <100> crystal axes which are invariant in symmetry to 90° rotations about them, then

$$C_{iiii} = C_{11}, C_{iijj} = C_{12} \quad \text{for i} \neq \text{j}$$
$$C_{ijij} = C_{44} \quad \text{for i} \neq \text{j}$$

and all other constants are zero because the sign of ϵ_{ij} for $i \neq j$ cannot affect the value of σ_{ii} for $C_{iiij} = 0$, etc. Thus, for cubic crystals the 6×6 elastic constant matrix in Eq. 30 can be written as

$$C_{mn} = \begin{bmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{bmatrix} \tag{31}$$

From Eq. 30, the relationships between stress, σ, and strain, ϵ, for cubic crystals can be written as

$$\begin{array}{ll} \textbf{Principal stresses and strains} & \left\{ \begin{array}{ll} \sigma_{11} & = C_{11}\epsilon_{11} + C_{12}\epsilon_{22} + C_{12}\epsilon_{33} \\ \sigma_{22} & = C_{12}\epsilon_{11} + C_{11}\epsilon_{22} + C_{12}\epsilon_{33} \\ \sigma_{33} & = C_{12}\epsilon_{11} + C_{12}\epsilon_{22} + C_{11}\epsilon_{33} \end{array} \right\} \\ \textbf{Shear stresses and strains} & \left\{ \begin{array}{ll} \sigma_{23} & = 2C_{44}\epsilon_{23} \\ \sigma_{31} & = 2C_{44}\epsilon_{23} \\ \sigma_{12} & = 2C_{44}\epsilon_{12} \end{array} \right\} \end{array} \tag{32}$$

Thus, we see that for cubic symmetry, there are three independent elastic constants: $C_{11,}$, C_{12}, C_{44}.

For isotropic cubic crystals the compliance tensor, S_{mn}, will appear in the form of Eq. 31, where

$$(C_{11} - C_{12})/2 = C_{44} = 1/S_{44} = G \tag{33}$$

and G is the shear modulus also expressed by

$$G = 1/2(S_{11} - S_{12}) \tag{34}$$

and the corresponding Young's modulus or elastic modulus is

$$E = 1/S_{11} = C_{11}. \tag{35}$$

Poisson's ratio is also expressed by

$$\upsilon = -S_{12}/S_{11}. \tag{36}$$

As it turns out, even polycrystalline aggregates can be expressed by Eq. 33, while it does not apply to cubic crystals explicitly unless they are perfectly isotropic. For anisotropic cubic crystals we define an anisotropy ratioas

$$A = 2C_{44}/(C_{11} - C_{12}), \tag{37}$$

Table 2 Anisotropic constants for cubic crystals at room temperature[a]

	Stiffness (×10 GPa)			Anisotropy ratio	Compliance (10^{-2}GPa^{-1})			Shear modulus	Poisson's ratio
Crystal	C_{11}	C_{12}	C_{44}	A	S_{11}	S_{12}	S_{44}	G × 10 GPa	υ
Al	10.82	6.13	2.85	1.21	1.57	0.568	3.51	2.65	0.347
Ag	12.40	9.34	4.61	3.01	2.29	0.983	2.17	3.38	0.354
Au	18.6	15.7	4.20	2.9	2.33	1.065	2.38	3.10	0.412
Cr	35.0	5.78	10.1	0.69	0.300	0.043	0.992	12.1	0.13
Cu	16.84	12.14	7.54	3.21	1.498	0.629	1.326	5.46	0.324
Fe	24.2	14.65	11.2	2.36	0.760	0.287	0.892	8.6	0.291
Ge	12.89	4.83	6.71	1.66	0.978	0.266	1.490	5.64	0.200
K	0.457	0.374	0.263	6.35	82.3	37.0	38.0	0.174	0.312
Mo	46	17.6	11.0	0.775	0.28	0.078	0.91	12.3	0.305
Na	0.603	0.459	0.586	8.15	48.6	21.0	17.1	0.380	0.201
Nb	24.6	13.4	2.87	0.51	0.660	0.233	3.48	3.96	0.392
Ni	24.65	14.73	12.47	2.52	0.734	0.274	0.802	9.47	0.276
Pb	4.66	3.92	1.44	3.90	9.28	4.24	6.94	1.01	0.387
Ta	26.7	16.1	8.25	1.56	0.686	0.258	1.212	7.07	0.339
Th	7.53	4.89	4.78	3.62	2.72	1.07	2.09	3.40	0.254
Si	16.57	6.39	7.96	1.74	0.767	0.214	1.256	6.41	0.215
V	22.8	11.9	4.26	0.78	0.683	0.234	2.348	4.73	0.352
W	52.1	20.1	16.0	1.00	0.244	0.068	0.624	16.0	0.278
AgBr	5.63	3.3	0.720	0.618	3.13	1.17	13.9	0.87	0.401
KCl	3.98	0.62	0.625	0.372	2.62	0.35	16.0	1.045	0.250
LiF	11.12	4.20	6.28	1.82	1.135	0.31	1.59	5.15	0.187
MgO	28.6	8.7	14.8	1.49	0.408	0.095	0.676	12.9	0.173
NaCl	4.87	1.24	1.26	0.694	2.29	0.465	7.94	1.48	0.248
PbS	12.7	2.98	2.48	0.614	0.864	0.164	4.03	3.23	0.269
Diamond	107.6	12.5	57.6	1.21	0.095	0.0099	0.174	53.6	0.068

[a]Data from Hirth and Lothe (1968)

Table 3 Elastic constants (measured at room temperature) for polycrystalline, cubic metals[a]

Metal (20 °C)	E (GPa)	G (GPa)	υ
Aluminum	70.3	26.1	0.345
Chromium	279.1	115.4	0.210
Copper	129.8	48.3	0.343
Gold	78.0	27.0	0.440
Iron	211.4	81.6	0.293
Lead	16.0	5.5	0.450
Nickel	199.5	76.0	0.312
Niobium	104.9	37.5	0.397
Silver	82.7	30.3	0.367
Tantalum	185.7	69.2	0.342
Tungsten	411.0	160.6	0.280
Vanadium	127.6	46.7	0.365

[a]Data from Hertzberg (1976)

where $A = 1$ for perfect isotropic behavior. Deviations from isotropic behavior can be expressed by $A > 1$ and $A < 1$. For isotropic solids, there are only two independent elastic constants: C_{11}, C_{12}.

It can be observed in Table 2 that some crystalline materials such as aluminum, diamond, molybdenum, and vanadium are reasonably isotropic ($A \sim 1$), while tungsten is the only isotropic material ($A = 1$). Sodium and potassium are extremely anisotropic. It is also noted in Table 2 that Eqs. 33, 35, and 36 are only applicable for tungsten since it is the only truly isotropic crystal. The values for the shear moduli, G, and Poisson's ratios listed in Table 2 are calculated from the so-called Voigt average (Voigt 1928):

$$G = \left(3C_{ijij} - C_{iijj}\right) \tag{38}$$

and

$$\upsilon = \frac{E - 2G}{2G}, \tag{39}$$

respectively. In Eq. 38, the local strain is equal to the mean strain in a polycrystalline aggregate where all grains are assumed to undergo the same strain. These values can be composed with measured (average) values for the elastic and shear moduli and Poisson's ratio, for polycrystalline cubic metals shown in Table 3.

As noted above the tensor values (Table 2) are only applicable for W, although the elastic modulus, $E = C_{11}$, even for tungsten is different in Table 3, and as a consequence, the tensor notation as developed in Eqs. 31 and 32 is not applicable to anisotropic (and therefore the vast majority of) cubic crystals and other structure materials. They simply provide a conceptual framework for materials properties or property coefficients.

References

Hertzberg RW (1976) Deformation and fracture mechanics of engineering materials. Wiley, New York, p 8
Hirth JP, Lothe J (1968) Theory of dislocation. McGraw-Hill, New York
Nye JF (1959) Physical properties of crystals : their representation by tensor and matrices. Oxford University Press, Oxford, UK
Voigt W (1928) Lehrbuch der Kristalphysik. Teubner, Leipzing

Tensor Transformation and Crystal Orientation Effects on Properties

Contents

Abstract

A novel matrix can execute a transformation from one set of Cartesian axes to another by systematic rotations through angles α, β, and γ or Euler angles: ϕ, θ, and ψ. In effect, this allows tensor coefficients representing a crystal property measured in a specific orientation to be mathematically transformed into any other orientation, thereby eliminating the necessity to remeasure the property coefficients.

Introduction

In the case of a second-rank tensor representing properties, such as resistivity illustrated in ▶ Eqs. 3 and ▶ 4 of chapter "Examples of Tensor Properties Using Matrix Fundamentals (A Physical Property)" in the context of the Cartesian coordinate system of axes shown in ▶ Fig. 2 (chapter "Examples of Tensor Properties Using Matrix Fundamentals (A Physical Property)"), the vector property such as E_i in ▶ Eq. 5 of chapter "Examples of Tensor Properties Using Matrix Fundamentals (A Physical Property)" can be evaluated along any direction relative to the axes, x, y, $z = x_1, x_2, x_3$, using a transformation formalism for either a vector or the coordinates of a point. In addition this can be done in the context of a crystallographic direction $[hkl]'$ relative to the directions [100], [010], and [001] for cubic symmetry which are characteristic of the principal components, $E_i = E_1, E_2, E_3$, for example.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_20

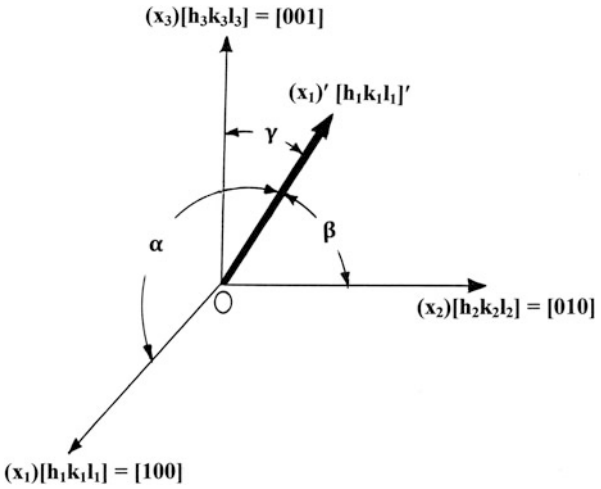


Fig. 1 Vector (direction) $[h_1k_1l_1]'$ or x_1' relative to the cubic or orthogonal principal axes (or directions); [100], [010], [001]

This is illustrated in Fig. 1, where $E_i' = E[hkl]' = E_p$. It can be seen that the relationship between the vector $E[hkl]'$ and the original directional properties $E[hkl] = E[100]$, $E[010]$, and $E[001]$ or vectors in the direction $x_p(x_1, x_2, x_3)$ and resolving in direction x_i (where $i = 1, 2, 3$) to give

$$E_i' = E_1 \cos \angle x_1 x_1' + E_2 \cos \angle x_2 x_1' + E_3 \cos \angle x_3 x_1', \tag{1}$$

where $\cos \angle x_1 x_1'$ is the cosine of the angle between x_i ($i = 1, 2, 3$), the original axes, and the axis (or new direction), x_1' or $[h_1k_1l_1]'$. The superscript is used to designate a new axis from the "old" or original axis: x_1 (old), x_1' (new).

Using this notation, and direction cosines between a complete set of axes, x_i, a transformation tensor table can be constructed showing the complete transformation of old axes, x_i to new axes, x_1', using direction cosines given by l_{ip}:

		Old		
		x_1	x_2	x_3
New	x_1'	l_{11}	l_{12}	l_{13}
	x_2'	l_{21}	l_{22}	l_{23}
	x_3'	l_{31}	l_{32}	l_{33}

(2)

		Old		
		$[h_1k_1l_1]$	$[h_2k_2l_2]$	$[h_3k_3l_3]$
New	$[h_1k_1l_1]'$	l_{11}	l_{12}	l_{13}
	$[h_2k_2l_2]'$	l_{21}	l_{22}	l_{23}
	$[h_3k_3l_3]'$	l_{31}	l_{32}	l_{33}

(3)

where Eq. 1 can be written as

$$E_i' = l_{i1}E_1 + l_{i2}E_2 + l_{i3}E_3, \tag{4}$$

$$E_i' = l_{ip}E_p. \tag{5}$$

Correspondingly, in terms of the crystal directions [hkl], Eq. 4 can be written as

$$E[h_1k_1l_1]' = l_{i1}E[100] + l_{i2}E[010] + l_{i3}E[001] \tag{6}$$

From this operation it is also possible to write

$$x_1{}' = l_{11}x_1 + l_{12}x_2 + l_{13}x_3 \tag{7}$$

or

$$x_i' = l_{ip}x_p \tag{8}$$

which represents the coordinates of the point represented by the vector axis x_1 or x_i (Eq. 8). Correspondingly, for the electric field components of the new (or primed) set of axes relative to the old axes as shown in Fig. 2,

$$E_i' = l_{ip}E_p = l_{ip}\,\rho_{pq}J_q \tag{9}$$

Similarly, the current density can be written relative to the new axes as

$$J_q = l_{jq}J_j', \tag{10}$$

Substituting Eq. 10 into Eq. 9

$$\begin{aligned} E_i' &= l_{ip}l_{iq}\rho_{pq}J_q \\ &= \rho_{ij}{}'J_j' \end{aligned} \tag{11}$$

and

$$\rho_{ij}{}' = l_{ip}l_{iq}\rho_{pq} \tag{12}$$

which illustrates the transformation of the original resistivity tensor, ρ_{pq}, in ▶ Eq. 4 of chapter "Examples of Tensor Properties Using Matrix Fundamentals (A Physical Property)" to the new resistivity tensor, $\rho_{ij}{}'$. This can be expressed generally for a second-rank tensor, T_{pq}, to be transformed to $T_{ij}{}'$:

$$T_{ij}{}' = l_{ip}l_{iq}T_{pq} \tag{13}$$

Equation 3 can also be expressed using a matrix transpose in the form:

$$T_{ij}' = l_{ij}T_{ij}l_{ij}^{T} = l_{ij}l_{ij}^{T}T_{ij} \tag{14}$$

where $l_{ij}^{T} = l_{ji}$.

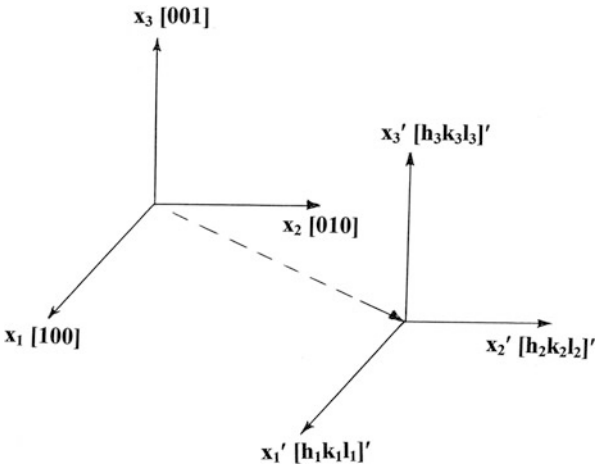


Fig. 2 Transformation of principal coordinates. [100], [010], [001] to $[h_1k_1l_1]'$, $[h_2k_2l_2]'$, $[h_3k_3l_3]'$, or x_1, x_2, x_3, $\rightarrow$ x_1', x_2', x_3'

Euler Angles and Transformation Examples

The transformation of axes, or orthogonal crystal directions illustrated in Fig. 2, can also be described in a series of angular rotation sequences referred to as *Eulerian (or Euler) angles* as illustrated in Fig. 2. Here, the new axes are represented by X, Y, Z in contrast to the original (old) axes x, y, z for the two coordinate systems. N in Fig. 3 is referred to as the line of nodes and is perpendicular to both the xy and XY coordinate planes and is also perpendicular to both z and Z. The Euler angles are given as α, β, and γ or alternatively ϕ, θ, and ψ; where α is the angle between the x-axis and N, β is the angle between z and Z, while γ is the angle between N and the X-axis. Therefore, α rotates about z, β rotates around N (or x′), and γ rotates about Z.

Figures 2 and 3, as they represent the transformation from a set of axes to another, especially for an isotropic cubic crystal, or any orthogonal crystals, illustrate a formal way for obtaining tensor components in a new orientation using those originally measured in a conventional orientation. In addition, the tensor coefficients can also be used to calculate magnitudes of different physical properties in different crystal directions, $[hkl]'$. This also means that coefficients of tensor properties measured for orthogonal crystals along the principal directions, x_1, x_2, x_3 or [100], [010], [001], can be determined for other orthogonal orientations without having to remeasure the vector quantities along the new principal axes (or directions).

As an example of tensor transformation, consider the recent determination of the thermal conductivity tensor for $Sr_4Bi_4Ti_7O_{24}$, by Zubuchen et al. (2012).

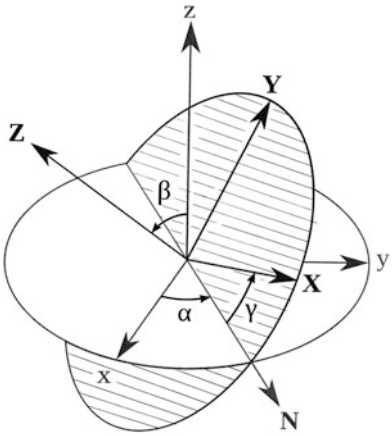


Fig. 3 Transformation of principal (orthogonal) axes x, y, z, to X, Y, Z using Euler angles: $\alpha, \beta, \gamma = \phi, \theta, \psi$

$Sr_4Bi_4Ti_7O_{24}$ is a layered perovskite phase and a member of the Aurivillius family of phases. These bismuth-layered oxides consist of $Bi_2O_2^{2+}$ layers alternating with 7 ABO_3 perovskite units. This is a very unusual tetragonal unit cell having a = b = 3.88 Å, c = 64.7 Å. The corresponding thermal conductivity tensor was determined to be

$$K_{ij} = \begin{pmatrix} K_{11} & 0 & 0 \\ 0 & K_{11} & 0 \\ 0 & 0 & K_{33} \end{pmatrix} = \begin{pmatrix} 1.8 & 0 & 0 \\ 0 & 1.8 & 0 \\ 0 & 0 & 1.1 \end{pmatrix} Wm^{-1}K^{-1}, \tag{15}$$

where in an isotropic medium,

$$h_i = K_{ij}\partial T/\partial x_j, \tag{16}$$

The h_i components represent the heat conduction along the principal axes (or directions) in ▶ Fig. 5 of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements," while $\partial T/\partial x_j$ represents the corresponding temperature gradient (°K/m).

If the original axes are rotated around the z-axis or the x_3 [001] axis by 45°, the new x_i axis rotated 45° counterclockwise as shown in Fig. 4 will become the [110] direction while the correspondingly rotated x_2 axis (x_2') will become the $[\bar{1}10]$ direction. The transformation tensor table (Eq. 2) can then be expressed by

		Old		
		$x_1[100]$	$x_1[010]$	$x_1[001]$
New	$x_1'[100]$	$\sqrt{2}/2$	$\sqrt{2}/2$	0
	$x_1'[110]$	$-\sqrt{2}/2$	$\sqrt{2}/2$	0
	$x_1'[001]$	0	0	1

(17)

From Eq. 3

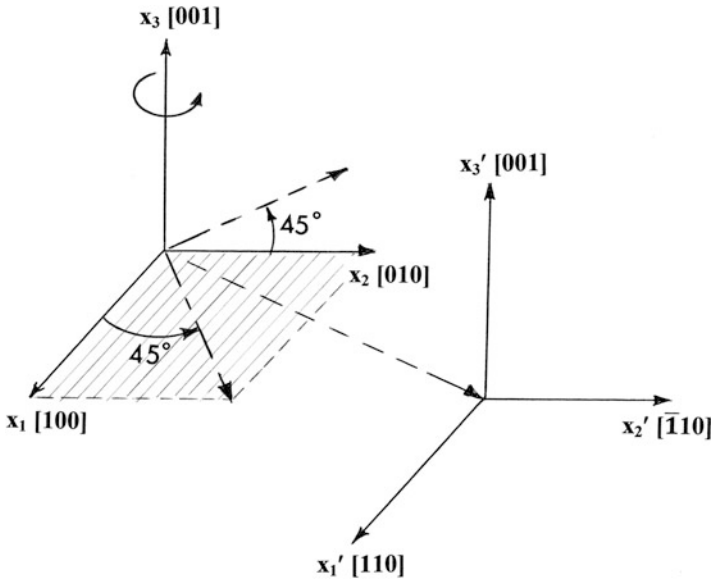


Fig. 4 Transformation of principal coordinates: [100], [010], [001] to [110], $[\bar{1}10]$, [001] by a rotation of 45° counterclockwise about [001] or x_3 (z-axis in Cartesian coordinates)

$$K'_{ij} = \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1.8 & 0 & 0 \\ 0 & 1.8 & 0 \\ 0 & 0 & 1.1 \end{pmatrix}$$
$$= \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1.8 & 0 & 0 \\ 0 & 1.8 & 0 \\ 0 & 0 & 1.1 \end{pmatrix} \tag{18}$$

or from Eq. 13 or 14:

$$K'_{11} = \ell_{1p}\ell_{1q}K_{pq}, \text{since } i = j = 1; p, q = 1, 2, 3$$

therefore

$$\begin{aligned} K'_{11} &= \ell_{11}\ell_{11}K_{11} = 1.8/2\,(p = q = 1); \ell_{ij} \Rightarrow \text{Eq. 2} \\ &+ \ell_{11}\ell_{12}K_{12} = 1.8/2(p = 1, q = 2) \\ &+ \ell_{11}\ell_{13}K_{13} = 0(p = 1, q = 3) \\ &+ \ell_{12}\ell_{11}K_{21} = 0(p = 2, q = 1) \\ &+ \ell_{12}\ell_{12}K_{22} = 1.8/2(p = 2, q = 2) \\ &+ \ell_{12}\ell_{13}K_{23} = 0(p = 2, q = 3) \\ &+ \ell_{13}\ell_{11}K_{31} = 0(p = 3, q = 1) \\ &+ \ell_{13}\ell_{12}K_{32} = 0(p = 3, q = 2) \\ &+ \ell_{13}\ell_{13}K_{33} = 0(p = 3, q = 3) \\ K'_{11} &= \overline{2.7}\mathrm{Wm}^{-1}\mathrm{K}^{-1} \end{aligned}$$

Continuing, it is observed that

$$K'_{12} = \ell_{1p}\ell_{2q}K_{pq}, \text{since } i = 1, j = 2; p, q = 1, 2, 3$$

therefore

$$\begin{aligned} K'_{12} &= \ell_{11}\ell_{21}K_{11} = -1.8/2\,(p = 1,\ q = 21) \\ &+ \ell_{11}\ell_{22}K_{12} = 1.8/2(p = 1, q = 2) \\ &+ \ell_{11}\ell_{23}K_{13} = 0(p = 1, q = 3) \\ &+ \ell_{12}\ell_{21}K_{21} = 0(p = 2, q = 1) \\ &+ \ell_{12}\ell_{22}K_{22} = +1.8/2(p = 2, q = 2) \\ &+ \ell_{12}\ell_{23}K_{23} = 0(p = 2, q = 3) \\ &+ \ell_{13}\ell_{21}K_{31} = 0(p = 3, q = 1) \\ &+ \ell_{13}\ell_{22}K_{32} = 0(p = 3, q = 2) \\ &+ \ell_{13}\ell_{23}K_{33} = 0(p = 3, q = 3) \\ K'_{12} &= \overline{0} \end{aligned}$$

Continuing, it can be observed that since the off-diagonal coefficients of K_{ij} (Eq. 16) are zero, K_{ij} will only have diagonal coefficients, so K_{13}, K_{23}, K_{33}, K_{31}, K_{32} will be zero:

$$K_{22}{}' = K_{11}{}' = 2.7 Wm^{-1}K^{-1}$$

and

$$K_{33}{}' = K_{33} = 1.1 Wm^{-1}K^{-1}$$

and the transformed tensor becomes

$$K_{ij}{}' = \begin{pmatrix} 2.7 & 0 & 0 \\ 0 & 2.7 & 0 \\ 0 & 0 & 1.1 \end{pmatrix} Wm^{-1}K^{-1} \tag{19}$$

There are several issues regarding the execution of a tensor transformation which are useful. First, it should be recognized that the transformation tensor table (in Eq. 18) can be checked by noting that

$$\sum \left(\ell_{ij}\right)^2 = 1; \tag{20}$$

$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$.

Therefore, for $x'_1 \Rightarrow \left(\sqrt{2}/2\right)^2 + \left(\sqrt{2}/2\right)^2 = 1$, etc. In addition, considering the crystal directions, $[h_i k_i l_i]$ and $[h_i k_i l_i]'$, the direction cosines are

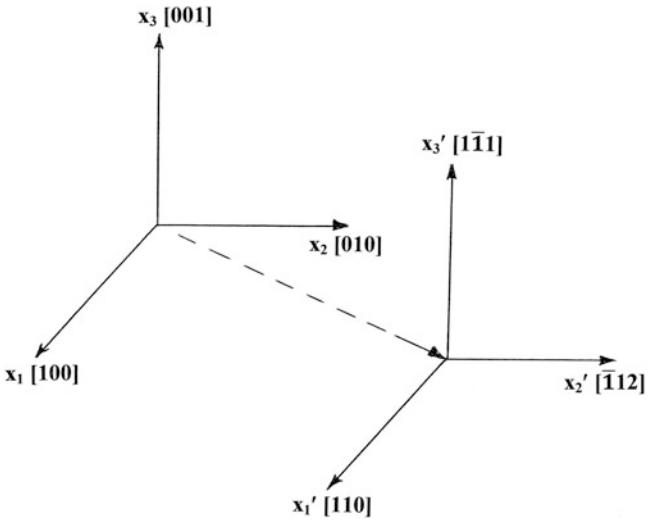


Fig. 5 Transformation of principal coordinates: [100], [010], [001] to [110], [$\bar{1}$12], [1$\bar{1}$1]

$$\text{Cos } \alpha = \frac{h_1h_1 + k_1k_1 + l_1l_1}{\sqrt{h_1^2 + k_1^2 + l_1^2}\sqrt{h_1'^2 + k_1'^2 + l_1'^2}} \equiv [h_1k_1l_1] \cdot [h_1k_1l_1]'$$

$$\text{Cos } \beta = \frac{h_2h_1 + k_2k_1 + l_2l_1}{\sqrt{h_2^2 + k_2^2 + l_2^2}\sqrt{h_1'^2 + k_1'^2 + l_1'^2}} \equiv [h_2k_2l_2] \cdot [h_1k_1l_1]'$$

$$\text{Cos } \gamma = \frac{h_3h_1 + k_3k_1 + l_3l_1}{\sqrt{h_3^2 + k_3^2 + l_3^2}\sqrt{h_1'^2 + k_1'^2 + l_1'^2}} \equiv [h_3k_3l_3] \cdot [h_1k_1l_1]'$$

Additionally, the orthogonality of the transformed crystal (or coordinate system) can also be checked by considering:

$$\left.\begin{array}{l} [h_1k_1l_1]' \cdot [h_2k_2l_2]' = 0 \\ [h_2k_2l_2]' \cdot [h_3k_3l_3]' = 0 \\ [h_1k_1l_1]' \cdot [h_3k_3l_3]' = 0 \\ [h_2k_2l_2]' \cdot [h_3k_3l_3]' = 0 \end{array}\right\} \tag{21}$$

representing that the axial directions, x_1', x_2', and x_3', are mutually perpendicular (90°) (Fig. 2).

As an example, one might consider the transformation illustrated in Fig. 5. Here, the corresponding transformation tensor and tensor table can be expressed as

$$
\begin{array}{c|ccc}
 & x_1[100] & x_1[010] & x_1[001] \\
\hline
x_1'[100] & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\
x_1'[110] & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\
x_1'[001] & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}
\end{array}
= l_{ij} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}
$$

It should be noted, of course, that tensor (and matrix) applications are available through *Mathcad* and other mathematics-related personal computer software. Nonetheless, this presentation has provided fundamental concepts which illustrate the utility of tensor-related tools as they apply to crystal structures and especially physical properties applied to crystal structures.

Although, as emphasized repeatedly, many materials can be treated as isotropic, they are often not so microscopically. However, in polycrystalline materials, individual grains exhibit crystalline anisotropy and symmetry (Kocks et al. 2000), but in aggregate form, and randomly oriented, the material can be considered microscopically isotropic, and the elastic constants are the same in all directions. In many cases, the elastic moduli (E) are different along three perpendicular directions, $E_1 \neq E_2 \neq E_3$, and this condition is referred to as orthotropic.

In cubic material, elastic and shear moduli (E and G) can be determined in any direction (or crystal orientation), using the elastic constants in the following equations :

$$1/E_{hkl} \equiv 1/E_{ijk} = S_{11} - 2(S_{11} - S_{12} - S_{44}/2)\left(f\left(l_{i1}, l_{j2}, l_{k3}\right)\right) \tag{22}$$

and

$$1/G_{hkl} \equiv 1/G_{ijk} = S_{44} + 4(S_{11} - S_{12} - S_{44}/2)\left(f\left(l_{i1}, l_{j2}, l_{k3}\right)\right) \tag{23}$$

where l_{i1}, l_{j2}, l_{k3} are direction cosines for the direction [ijk] ≡ [hkl] and the function

$$f\left(l_{i1}, l_{j2}, l_{k3}\right) = \left(l_{i1}^2 l_{j2}^2 + l_{j2}^2 l_{k3}^2 + l_{il}^2 l_{k3}^2\right). \tag{24}$$

This is illustrated in Fig. 1 for a direction $[h_1k_1l_1]' = [hkl]$ in a cubic crystal lattice unit cell, or in a transformed crystal notation, we can use the following:

$$
\begin{array}{c|ccc}
 & [100] & [010] & [001] \\
\hline
[hkl] & l_{i1} & l_{i2} & l_{k3} \\
ijk & & &
\end{array}
\quad - \quad
\frac{\left(l_{i1}^2 l_{j2}^2 + l_{j2}^2 l_{k3}^2 + l_{i1}^2 l_{k3}^2\right)}{(\,)}
$$

For example, for the [100] direction,

Table 1 Calculated elastic moduli for some common cubic metals

Metal	Elastic modulus (GPa)		
	E_{100}	E_{110}	E_{111}
Aluminum (fcc)	64	72	76
Nickel (fcc)	136	233	303
Copper (fcc)	66	130	191
Iron (bcc)	125	210	272
Molybdenum (bcc)	360	305	292
Tungsten (bcc)	390	390	390

$$\begin{array}{c|ccc} & [100] & [010] & [001] \\ \hline [100] & 1 & 1 & 1 \end{array} \quad - \quad \frac{\left(l_{i1}^2 l_{j2}^2 + l_{j2}^2 l_{k3}^2 + l_{i1}^2 l_{k3}^2\right)}{(0)}$$

or for [110]:

$$\begin{array}{c|ccc} & [100] & [010] & [001] \\ \hline [110] & \sqrt{2}/2 & \sqrt{2}/2 & 0 \end{array} \quad - \quad \frac{\left(l_{i1}^2 l_{j2}^2 + l_{j2}^2 l_{k3}^2 + l_{i1}^2 l_{k3}^2\right)}{(1/4)},$$

and correspondingly, $[111] \Rightarrow (l_{i1}^2 l_{j2}^2 + l_{j2}^2 l_{k3}^2 + l_{i1}^2 l_{k3}^2) = (1/3)$. Consequently, one can calculate E_{100}, E_{110}, and E_{111} for cubic metals where the compliances, S_{11}, S_{12}, and S_{44}, are known (▶ Table 2 of chapter "Examples of Tensor Properties Using Matrix Fundamentals (A Physical Property)") and using Eq. 22. Calculated values for a few common cubic metals using Eq. 22 are listed in Table 1.

It is of interest to examine the trends of the elastic moduli in Table 1 which, in comparison to elastic constants in ▶ Table 2 (chapter "Examples of Tensor Properties Using Matrix Fundamentals (A Physical Property)"), are related in some respect to the anisotropy ratio, A (▶ Eq. 37 of chapter "Examples of Tensor Properties Using Matrix Fundamentals (A Physical Property)"), as well as the variance in specific elastic constants. This is especially notable for Mo in contrast to Fe, as well as the comparison of fcc Al and Cu. As noted previously (and in concert with ▶ Table 2 of chapter "Examples of Tensor Properties Using Matrix Fundamentals (A Physical Property)"), W is the only truly isotropic metal (A = 1). In the context of Table 1 and Eq. 22, it can be appreciated that where texture, especially strong texture as in directional solidification, and especially in growing single crystals of metals and particularly complex alloys (such as superalloy turbine blades, etc.), the applicable elastic modulus may be important. In addition and as implicit in even the limited and somewhat specific data in Table 1, assumptions regarding modulus trends may be completely unfounded, and a calculational diagnostic facilitated by Eq. 22 may be an important process and performance design strategy. This can also be an important feature of fiber (reinforcement)

composites where the fiber modulus and orientation have important consequences to be discussed later in connection with uniaxial (directional) fiber composites or laminated orthotropic composites.

References

Kocks UF, Tome CH, Werk H-R (2000) Texture and anisotropy preferred orientations in polycrystals and their effects on materials properties. Cambridge University Press, Cambridge, UK

Zurbuchen MA, Cahill DG, Schubent J, Jia Y, Schlom DG (2012) Determination of the thermal conductivity tensor of the n = 7 Aurivillius phase $Sr_4Bi_4Ti–O_{24}$. Appl Phys Lett 101: 021904-1 to 4

Elastic Anisotropy in Deformed (Textured) and Directionally Grown Crystalline and Polycrystalline Materials

Contents

Abstract

In this short chapter, the ability to characterize textured polycrystalline materials using an averaging technique for elastic moduli is briefly described.

Introduction

While single-crystal orientations can dictate properties, polycrystalline materials having component crystal grains oriented identically or nearly identically can have the same effects. Correspondingly, proportionate orientations representing texture can also influence properties somewhat in proportion to the proportion of the component orientations. In the case of elastic modulus or Young's modulus, the largest orientation grain fraction would be expected to dominate, even in the context of the average modulus characterizing texture. This is particularly true in metals and alloys where texture occurs in forming operations such as rolling or extrusion.

Texture Effects

In contrast to isotropic materials, anisotropic properties at a point in a material vary with direction or orientation of reference axes. If properties along any direction are the same as those along a symmetric direction with respect to a

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_21

plane, then that plane is referred to as a plane of material symmetry. Orthotropic materials have at least three mutually perpendicular axes called principal axes or principal materials axes. For orthotropic materials with their planes of symmetry aligned parallel to the coordinate axes as represented in ▶ Figs. 2 and ▶ 5 of chapter "Examples of Tensor Properties Using Matrix Fundamentals (A Physical Property)," Hooke's law can be expressed by the following compliance tensor:

$$\begin{pmatrix} \epsilon_x \\ \epsilon_y \\ \epsilon_z \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{xz} \end{pmatrix} = \begin{pmatrix} \frac{1}{E_1} & \frac{-\nu_{21}}{E_2} & \frac{\nu_{21}}{E_3} & 0 & 0 & 0 \\ \frac{-\nu_{21}}{E_1} & \frac{1}{E_1} & \frac{-\nu_{32}}{E_3} & 0 & 0 & 0 \\ \frac{-\nu_{13}}{E_1} & \frac{-\nu_{23}}{E_2} & \frac{1}{E_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{G_{12}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{G_{23}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{G_{13}} \end{pmatrix} \begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{xz} \end{pmatrix} \tag{1}$$

where principal axes 1, 2,3 coincide with the Cartesian axes x, y, z and the shear stresses, G_{ij}, are in planes parallel to 12, 13, 23 and ν_{ij} (Poisson's ratio) characterizes normal strains in the symmetry directions which coincide with stresses applied in these directions. In cubic crystals, Young's modulus in some arbitrary direction (or crystallographic direction) illustrated in ▶ Fig. 1 in chapter "Tensor Transformation and Crystal Orientation Effects on Properties," and expressed by Eq. 22, can be written as

$$\frac{1}{E_{hkl}} = \frac{1}{E_\theta} = \frac{1}{E_{11}} - 2\left[\frac{1}{E_{11}} + \frac{\nu_{12}}{E_{11}} - \frac{1}{2G}\right] f\left(l_{ij}, l_{j2}, l_{k3}\right) \tag{2}$$

where $f(l_{ij}, l_{j2}, l_{k2})$ is the direction cosine function given by ▶ Eq. 24 (chapter "Tensor Transformation and Crystal Orientation Effects on Properties"). Equation 2 is applicable to crystal directions or single-crystal rods or fibers in the same way ▶ Eq. 24 of chapter "Tensor Transformation and Crystal Orientation Effects on Properties" was applied.

Equation 2 can also be applied to polycrystalline regimes, where because of texturing or directional growth, some degree of anisotropy is created which, in the context of elastic moduli (Kocks et al. 2000), might be described by some average directional modulus in the form

$$\frac{1}{E(\text{ave.})} = \frac{1}{E_{11}} - 2\left[\frac{1}{E_{11}} + \frac{\nu_{12}}{E_{11}} - \frac{1}{G}\right] f(\theta)(\text{ave.}) \tag{3}$$

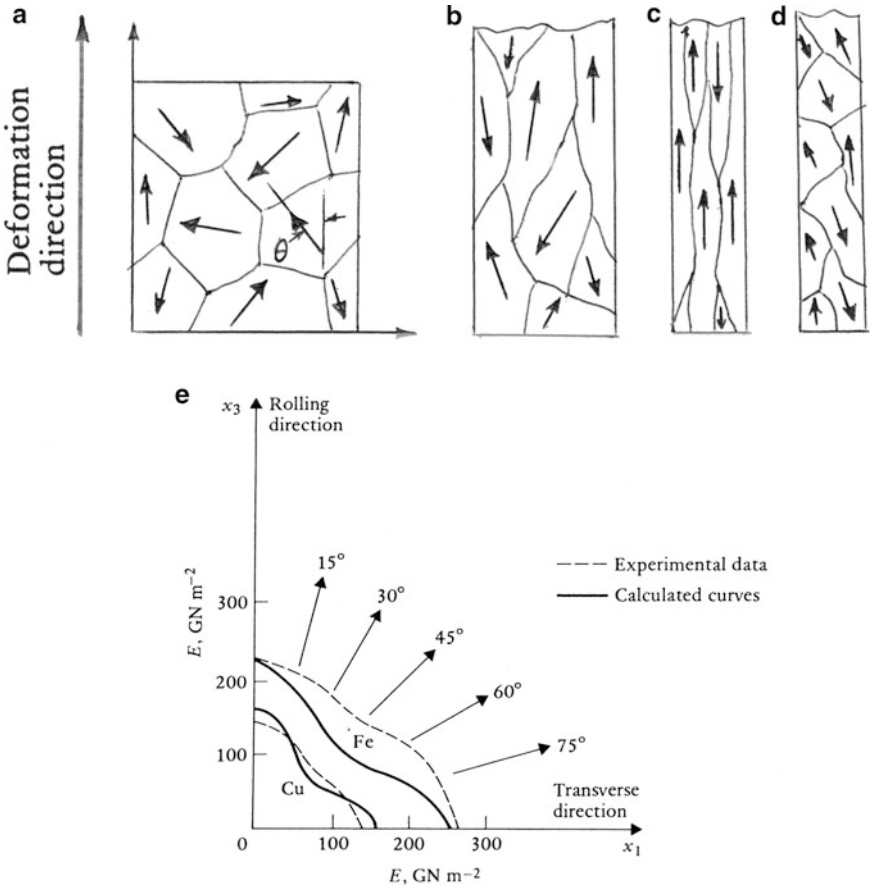


Fig. 1 Texture development (**a–d**) and variation of elastic modulus with orientation for Fe and Cu (**e**). (**a**) Polycrystalline section with grains oriented θ with respect to deformation axis (x_3). (**b**) and (**c**) show grain elongation and texture with deformation along x_3. (**d**) Random texture. (**e**) Variation of elastic modulus with orientation in the plane of cold-rolled sheets of Fe and Cu. Note orientation angles (θ) corresponding to (**a**). (**e**) is adapted from Harris and Bunsell (1977)

where f(θ) (ave.) is an averaged value of the direction cosine function

$$\left(l_{i1}^2 l_{j2}^2 + l_{j2}^2 l_{k3}^2 + l_{i1}^2 l_{k3}^2\right)(\text{ave.}) \tag{4}$$

In a random polycrystalline aggregate, the cube axes will lie between 0° and 90° to the reference axis, x_3, as illustrated in the schematic of Fig. 1a. Referred to ▶ Fig. 1 in chapter "Tensor Transformation and Crystal Orientation Effects on Properties," values for f(θ) (ave.) (Eq. 4) were 0 and 1/3 in the <001> and <111> directions, respectively, although a value of f(θ) (ave.) = 1/5 has been calculated for a random aggregate. For the case of wire drawing, extrusion, or rolling

deformation shown schematically in Fig. 1b and c, some fraction of the deformed and elongated grains can have a texture $<hkl>$ and a corresponding elastic (Young's) modulus, $E_{<hkl>}$. The average modulus can then be approximated from

$$E(\text{ave.}) = \sum_{\substack{<hkl>n \\ Vf<h_nk_nl_n>E_{<h_nk_nl_n>}}} V_F E_{<hkl>} = V_{f<h_1k_1l_1>} E_{<h_1k_1l_1>} + V_{f<h_2k_2l_2>} E_{<h_2k_2l_2>} + \ldots \quad (5)$$

where $V_{f<h_nk_nl_n>}$ is the volume fraction of directional grains having an orientation (or texture) $< h_nk_nl_n >$ and with a corresponding elastic modulus $E_{<h_nk_nl_n>}$.

As shown in Fig. 1d, some directional deformation produces random moduli, and the measured moduli in the rolling direction as well as the transverse directions (x_1 and x_2 in the sheet thickness) will be essentially the same. This is illustrated for iron and copper in Fig. 1e. It is interesting to note that in the case of Caribbean steel drum carbon steel sheet (▶ Fig. 3a, chapter "Point Defects"), the measured values of elastic constants, $E_{11} = E_1 = E_{22} = E_2$, $E_{33} = E_3$, in unrolled, steel sheet corresponding to the large rim note area were $E_1 = 211$ GPa, $E_2 = 209$ GPa, $E_3 = 210$ GPa (Ferreyra et al. 2001). These values correspond roughly to those for iron in Fig. 1e. In this regard, it can be noted from ▶ Eq. 6 in chapter "Point Defects" that the fundamental and harmonic frequencies depend upon the elastic modulus, E, and consequently, the fact that E is essentially constant provides a tonic reference for each note. In addition, it can be observed from ▶ Eq. 6 of chapter "Point Defects" that using a specific tonic or harmonic frequency, elastic (Young's) moduli can be measured from longitudinal sound velocities.

Of course knowing the directional elastic modulus for any material ($E_{<hkl>}$) can allow for the estimation of the applicable modulus in a textured material by using Eq. 5. The volume fraction of the specific texture (or textures) can be determined from orientation imaging microscopy such as electron backscatter diffraction (EBSD) analysis in the SEM or orientation image mapping by x-ray diffraction, especially using pole figure measurement software which is readily available (MATLAB tool box for texture analysis). These will allow estimates of the various texture fractions corresponding to volume fractions $V_{f<hkl>}$ in Eq. 5.

References

Ferreyra E, Murr LE, Russell DP, Bingert JF (2001) Elastic interactions and the metallurgical and acoustic effects of carbon in the Caribbean steel drum. Mater Charact 47:325–363

Harris B, Bunsell AR (1977) Structure and properties of engineering materials. Longman Group, London

Kocks UG, Tome CM, Wenk H-R (2000) Texture and anisotropy: preferred orientations in polycrystals and their effects on materials properties. Cambridge University Press, Cambridge, UK

Examples of Directional Crystal Structures: Gas-Turbine Component Applications in Superalloys

Contents

Abstract

One of the most notable applications of single-crystal orientation to control elastic modulus and associated mechanical properties is vested in the production and use of directionally grown turbine blades from superalloy compositions. This chapter provides examples of this phenomenon from both a historical and contemporary perspective. These examples epitomize the development and control of microstructure to achieve specific properties and performance based on processing schedules and selective routines.

Introduction

The development of high-performance, superalloys spanning a period of more than a half century epitomizes one of the more compelling examples of materials science and engineering, encompassing close coupling of structure/microstructure and properties controlled by processing routes to achieve specific performance criteria. Aircraft gas-turbine engine and especially requisite component materials development began just prior to World War II with precipitation-hardened and heat-treatable nickel-base alloys containing aluminum, notably Nimonic series alloys such as Nimonic 80 and later Nimonic 90. Early materials design strategies, which have been retained up to the present time, have included concerns for creep and creep-rupture resistance, fatigue and thermal fatigue, toughness, thermal shock

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_22

resistance, crack growth resistance, and oxidation resistance to corrosive, high-temperature, and high-velocity gas flow. Early turbine blades containing equiaxed grain boundaries were limited by rapid grain boundary diffusion and accelerated creep at higher temperatures, and directional solidification and growth processes were developed to produce columnar-grain structures which reduced diffusional creep, especially when grain misorientations were small. Pratt-Whitney Aircraft developed directional growth for single-crystal turbine blades in the 1960s, eliminating grain boundaries, although both polycrystalline and columnar-grain products continued to be used for less demanding turbine engine components. The greatest demand, which continues as a material development concern, is the high-temperature turbine section blade tolerance, especially sufficient creep strength maintenance to avoid airfoil lengthening. Contemporary blades used in the high-temperature turbine section are complex, hollow airfoil structures to reduce weight and facilitate cooling. These structures contain thickness variations and airfoil axis variations which create complex (multiaxial) stress distributions. In addition, and as represented in Table 1, nickel-base superalloys, like nickel, exhibit anisotropic elastic moduli. However, [001] has been the direction of choice even up to the present time since it provides the best combination of turbine hot-section properties, including the fact that a lower elastic modulus enhances thermal fatigue resistance. Correspondingly, temperature gradients in complex blade structures create thermal stresses in various directions.

Nickel-Base Superalloy Development Strategies for Gas Turbines

As noted above, early Nimonic as well as Inconel nickel-base alloys became the early cornerstones for gas-turbine jet engine metallurgy. This expanded to a wide range of continuing alloy development especially with the expansion of advanced gas-turbine engines for power generation and the need for advanced materials in nuclear power, chemical, and other processing plants. The particular challenge involves the continuing development of increased temperature requirements for small turbine blades of only a few centimeters in high-performance jet aircraft in contrast to land-based gas-turbine power generation systems having blade lengths more than an order of magnitude longer (50–80 cm). This became especially demanding in the investment casting processes used to fabricate elaborate airfoil cooling structures and simultaneous grain-structure control, especially the growth of oriented single crystals. Cooler stages of the turbine and other less demanding components continue to utilize equiaxed and columnar-grain structures.

Figure 1a illustrates a basic gas-turbine schematic view, while Fig. 1b shows the three common blade microstructures. Compressor blades utilize high yield and tensile strength, low density, high impact, and fatigue strength alloys, where temperatures range to ~600 °C. Inconel 718 has been a mainstay in compressor blade and vane applications up to ~700 °C. Inconel 625, 718 and Rene 88, 95 alloys are employed in turbine disc, shaft, fastener, and other engine hardware applications. Turbine rotor, hot-stage blades experience gas temperatures of ~1,200 °C at

Table 1 Compositions of some commercial Ni-base superalloys (weight %): columnar-grain space or single crystal except where noted. All alloys balance Ni

Alloy	Cr	Co	Al	Ta	W	Mo	Ti	Hf	C	B	Nb	Other
[a]Mar-M200	9	10	5	–	12.5	–	2	Note 1	0.15	–	–	1Cu, 0.5 Zr
[a]Nimonic 80A	19.5	–	1.4	–	–	–	2.4	–	0.06	0.003	–	0.06 Zr
[a]Rene′ 41	19	11	1.5	–	–	10	3.1	–	0.09	0.005	–	0.06Zr
[a]Rene′ 80	14	9.5	3	–	4	4	5	–	0.17	0.02	–	0.03Zr
CMSX-2	8	5	5.5	6	8	0.6	1	–	–	–	–	–
CMSX-3	8	7.5	5.6	6	8	0.6	1	0.1	–	–	–	–
CMSX-4	6.6	9.2	5.6	6.2	5.9	0.6	–	–	–	–	–	3.3 Re
Rene′ N6	4.2	12.5	5.8	7.2	6	1.4	–	0.15	0.05	–	–	4.5 Re
TMS-75	3	12	6	6	6	2	–	0.1	–	–	–	5 Re
MC2	8	5	5	6	8	2	1.5	–	–	–	–	–
CMSX-10	2	3	5.7	8	5	0.4	0.2	0.03	–	–	0.1	–
CMSX-10 M	2	1.75	5.8	8.2	5.4	0.4	–	–	–	–		–
MXON	8	5	6.1	6	8	2	–	–	–	–	–	–
Waspaloy 454	10	5	5	12	4	–	1.5	–	–	–	–	–
MC653	4	–	5.8	6	5	4	–	0.1	–	–	–	3 Re
MC-NG	4	–	6	5	5	1	0.5	0.1	–	–	–	4 Re

(*continued*)

Table 1 (continued)

Alloy	Cr	Co	Al	Ta	W	Mo	Ti	Hf	C	B	Nb	Other
AM1	7.8	6.5	5.2	7.9	5.7	2	1.1	–	–	–	–	–
MC-544	4	–	5.8	6	5	1	0.5	0.1	–	–	–	–
PWA 1480	10	5	5	12	4	–	1.5	–	–	–	–	–
PWA 1484	5	10	5.6	9	6	2	–	0.1	–	–	–	3 Re
[a]IN 718	19	–	0.5	–	–	3	0.9	–	–	0.02	5.1	18.5 Fe
[a]IN 625	22	–	0.2	–	–	9	0.3	–	0.02	–	3.5	3.2 Fe
[a]Mar-M247	8	9	5.5	–	9.5	–	–	–	–	–	–	–
[a]Mar-M246	8.3	10	5.5	3	10	0.7	1	1.5	0.14	0.02	–	0.05 Zr
[b]Rene′ 95	13	8	3.5	-	4	4	2.5	–	0.065	0.01	3.5	0.05 Zr
[c]GTD 111	13.5	9.5	3.3	2.7	3.8	1.5	4.8	–	0.09	0.01	–	0.23 Fe
[c]Rene′ 142												

Note 1: Mar-M200 (Hf) has a similar composition with 0.1 Hf
[a]Conventional casting or wrought product
[b]Powder processed
[c]Columnar-grained, directional solidification

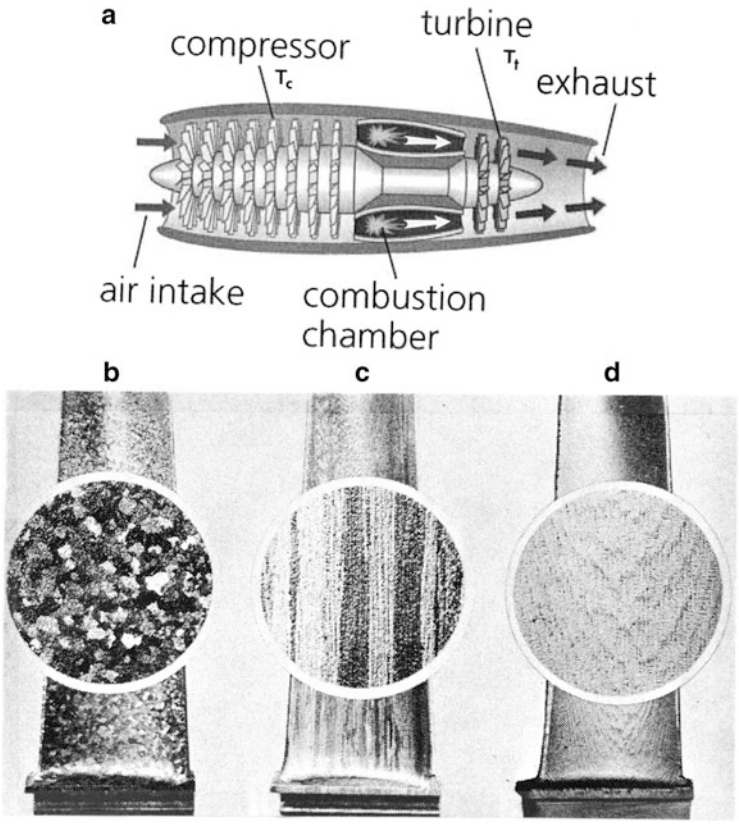


Fig. 1 (**a**) Simple gas-turbine schematic. Compressor and turbine regime temperatures are denoted T_c and T_t, respectively. (**b–d**) Comparative microstructures for three turbine blade (cast) types. (**b**) Polycrystalline; (**c**) columnar (directional) grained; (**d**) single crystal (Adapted from Oblak and Kear 1972)

first-stage rotors and lower temperatures at the exhaust section. Blade tip speeds average ~400 m/s and the initial gas velocity can exceed 600 m/s. Table 1 lists a wide range of Ni-base superalloy compositions which will serve as a reference for specific creep-related properties to be described in connection with continuing alloy development implicit in Fig. 2.

Two precipitation hardening features have characterized Ni-base superalloy development. These involve γ' (fcc) Ni_3 (Al, Ti) precipitation which controls properties of the Nimonic and progressive alloys and γ'' (bct) Ni_3 Nb precipitation which can be controlling in appropriately heat-treated Inconel alloys. These precipitation modes are morphologically distinct and have specific property and performance control features related to these morphologies. In each case, there is a variance in morphology and precipitate size controlled by heat treatment, which in some instances involves elaborate schedules. For the γ'' (bct) precipitates,

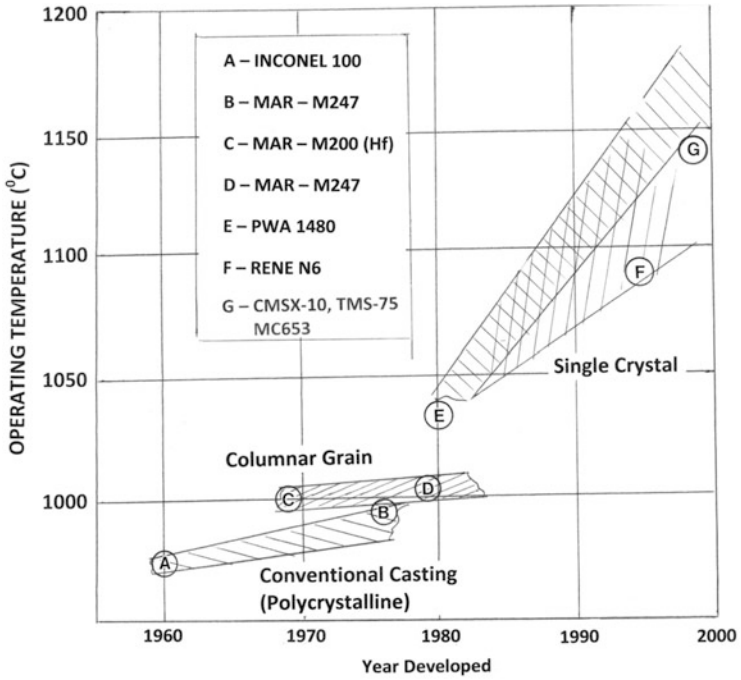


Fig. 2 Turbine blade Ni-base superalloy development examples. Double-hatched region shows proposed regime projected into the twenty-first century approaching 1,200 °C operating temperatures for high-performance gas-turbine operation

morphologies range from small nano-size, homogeneous spheres to discs or ellipsoidal platelets ($>100 \times 30$ nm) coincident with the γ (matrix), {100} planes. Correspondingly, the γ' (fcc) precipitates can also form as nano-size spheres which evolve into bimodal mixtures of small and larger, rounded cuboidal particles to large cuboidal particles, both coherent and coincident with γ(fcc) matrix {100} planes. In high-temperature turbine blade development, the focus has been on Ni-base superalloy, single-crystal development with γ'-cuboidal precipitate creep control.

At this point, it is probably expedient to review creep fundamentals. These are implicit in Fig. 3a which illustrates the three characteristic creep stages: primary, secondary, and tertiary creep. There is an initial, instantaneous elastic strain region similar to the $\sigma - \in$ diagram (▶ Fig. 9c in chapter "Line Defects: Dislocations in Crystalline Materials"). In the primary creep regime (stage I), $\in$ increases with time to the 1/3 power, while the strain rate ($\dot{\in}$) decreases because dislocation structures are changing to be in equilibrium with the applied stress, forming dislocation cell structures with increasing dislocation cell wall thickness and density as the cell size decreases. This phenomena is shown generally for primary creep in polycrystalline copper in Fig. 4. At the end of primary creep, dislocation cells continue to develop, especially in high stacking-fault energy metals, although

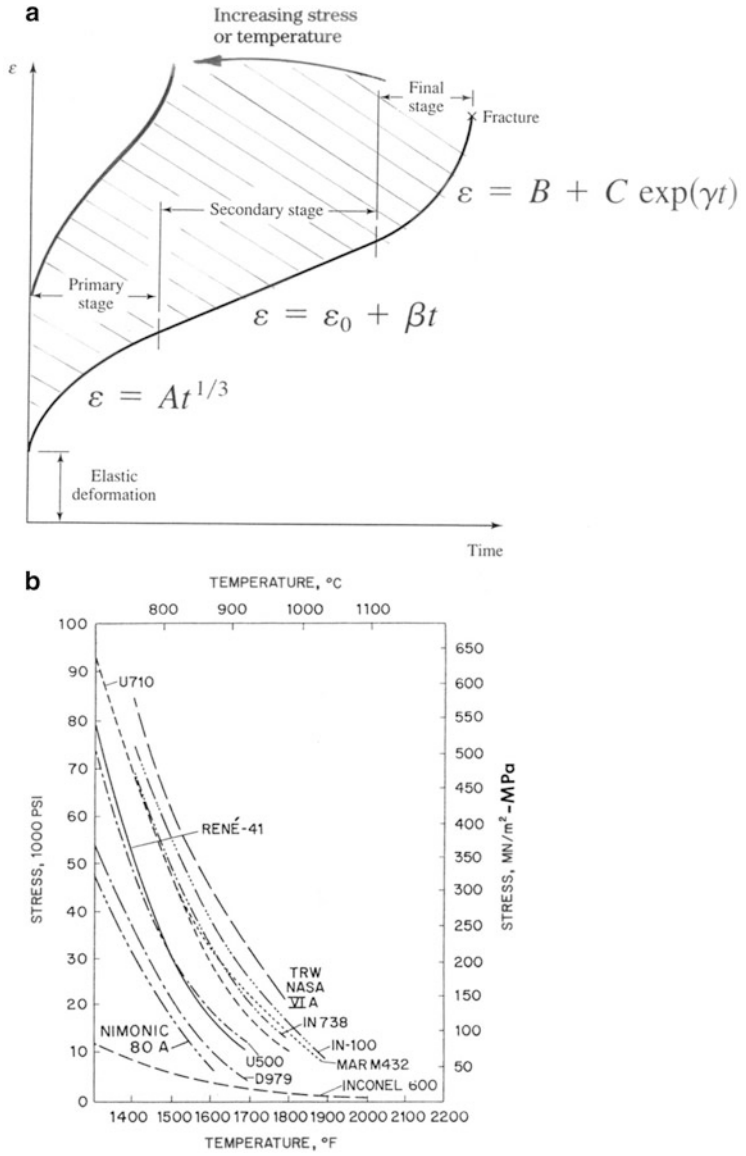


Fig. 3 Creep fundamentals (**a**) and examples of creep rupture (1,000 h) for stress versus temperature curves, for example, Ni-base superalloy compositions (Table 1) (**b**). Empirical representations are shown in (**a**) for three creep stages: primary (I), secondary (II), and tertiary (III) or final stage. (**b**) is adapted from Sims and Hagel (1972)

creep deformation is distinguishing, especially at the end of primary creep and into the secondary or steady-state regime where dislocation cell structure forms even for low stacking-fault energy metals and alloys. The distinguishing feature in creep deformation is the very regular dislocation cell structure in contrast to other

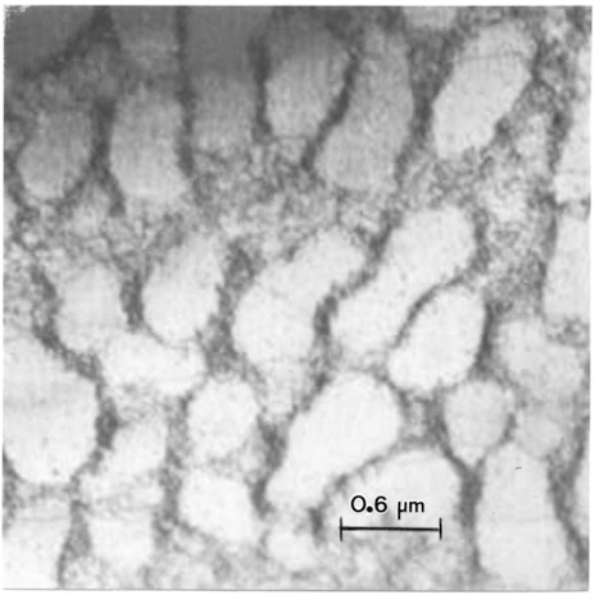


Fig. 4 Creep-related dislocation cell structure in copper near the end of primary (Stage (I)) creep

deformation modes illustrated on comparing Fig. 4 with ▶ Figs. 29a and ▶ 30 of chapter "Line Defects: Dislocations in Crystalline Materials." For copper, this dynamic equilibrium (or steady state) established between the stress (or creep load) and microstructure minimizes the creep rate in stage II (or the secondary creep regime) as expressed by the linear creep relation shown in Fig. 3a. At the end of the stage II linear creep regime, dislocation glide and interaction are aided by vacancy creation and diffusion which promotes dislocation climb, as well as some grain boundary gliding especially for small grain sizes and higher temperatures and stresses, all of which nucleate and grow cracks leading to failure. This stage accounts for most of the deformation which is effectively exponential in terms of the creep rate. Diffusion in the final creep state (III) becomes increasingly important with temperature increase, especially at $T > 0.3\ T_m$ (T_m = melting temperature). At $T\ 0.5\ T_m$, creep becomes dominant in many stress-related design considerations along with higher stresses as evident in Fig. 3a. Figure 3b shows some stress versus temperature curves for a range of nickel-base superalloys (Table 1) for creep-rupture life after 1,000 h (Fig. 3a).

Deformation Mechanism Maps: Creep Applications

Figure 5a and b illustrates a novel concept for synthesizing and visualizing a large body of complex deformation information (both theoretical and experimental) for materials in deformation maps originally conceived and extensively developed by M.F. Ashby (1972, 1976). These maps, as illustrated in Fig. 5a, b are particularly applicable to creep. A deformation mechanism map is constructed by using the rate equation (or shear strain rate equation) relating shear stress, temperature, and shear

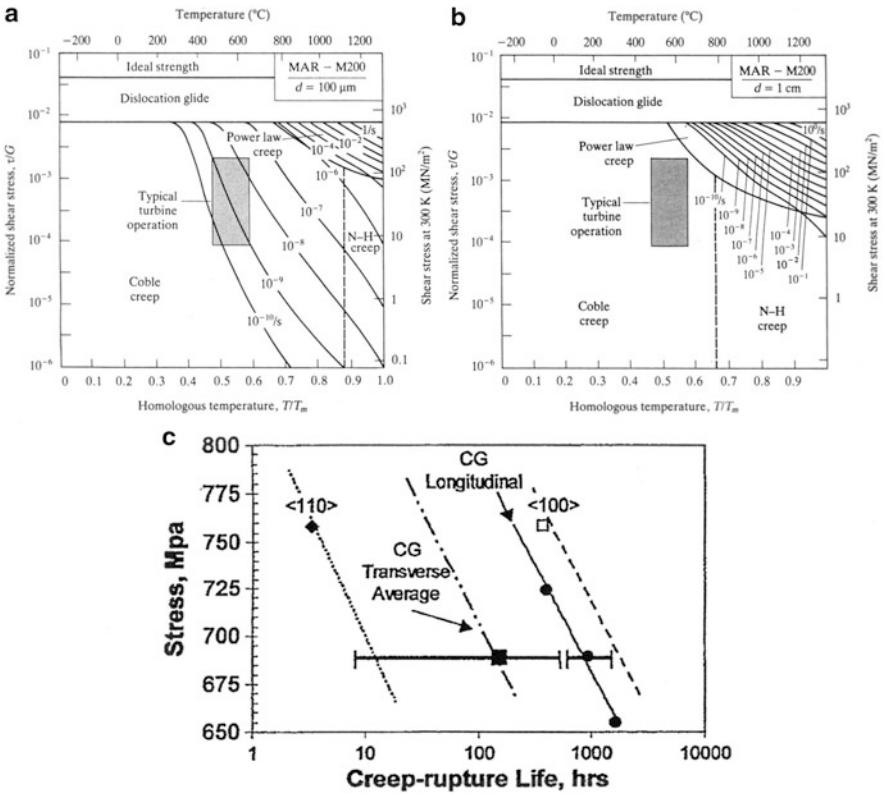


Fig. 5 (**a** and **b**) are examples of deformation mechanism maps for Mar-M200 Ni-base superalloy having grain sizes of 100um and 1 cm, respectively. The typical gas-turbine operating regime is shown by shaded rectangle (Adapted from Ashby (1973)). (**c**) Transverse 760 °C creep-rupture life for columnar grain (CG) and <100> and <110> single-crystal Mar-M200 (Hf) versus creep stress (Adapted from Shah and Cetel (2000)

strain rate for each deformation mechanism presented with axes of normalized shear stress (τ/G) and homologous or normalized temperature (T/T_m), as illustrated in Fig. 5a, b. The map is divided into fields indicating regions of stress and temperature where each deformation mechanism dominates. Contours of constant strain rate are superimposed on these fields.

There are four major deformation mechanisms and corresponding shear strain rates represented in the maps of Fig. 5a, b.

Dislocation Glide ($T < 0.3\ T_m$)

$$\dot{\gamma}_1 = \dot{\gamma}_0 \exp\left[-\frac{\Delta F}{\kappa T}\left(1 - \frac{\tau L}{Gb}\right)\right], \tag{1}$$

where $\dot{\gamma}_0$ is a constant (usually between 10^6 and 10^8 s^{-1}) and ΔF is the activation energy required to move a dislocation and has the form

$$\Delta F = \Delta F_\kappa \left(1 - \frac{\tau L}{Gb}\right)^2. \tag{2}$$

Here, ΔF_κ is the energy of formation of a dislocation kink pair, τ is the shear stress, G is the shear modulus, b is the Burgers vector, and L is the spacing of forest dislocations (or pinning points) or other obstacle spacing such as precipitate particles or noncoherent dispersed particles, often with a value $\sim 10^2 b$.

In addition to dislocation glide dependence on obstacles as described, the lattice itself can provide some resistance, particularly at low temperature, commonly referred to as the *Peierls resistance* where the strain rate is given by

$$\dot{\gamma}_2 = \dot{\gamma}_p \exp\left[-\frac{\Delta F}{\kappa T}\right] \tag{3}$$

where $\dot{\gamma}_p$ is typically 10^6–10^8 S^{-1} $\left(\dot{\gamma}_p \sim \dot{\gamma}_0\right)$.

Dislocation Creep ($T \geq 0.5T_m$)

As noted previously in describing creep (Fig. 13a), at temperatures above half the melting point, dislocations can climb (because of vacancy mobility) as well as glide, and deformation can occur at lower stresses than would be required for glide. The corresponding steady-state creep rate for moderate stresses and high temperatures is described by

$$\dot{\gamma}_3 = A \frac{D_{eff}\, Gb}{\kappa T} \left(\frac{\tau}{G}\right)^n, \tag{4}$$

where A and n are material constants, n having values, between 3 and 7 (Ashby 1976). D_{eff} is defined as an effective diffusivity:

$$D_{eff} = D_v V_v + D_c V_c \tag{5}$$

where D_v is the lattice (volume) diffusivity, V_v is the volume fraction of atom sites associated with D_v, and D_c is the dislocation core diffusivity having a volume fraction $V_c = A_c \rho$, where A_c is the dislocation core cross-sectional area and ρ is the dislocation density. Generally, $D_c = D_{gb}$, the grain boundary diffusivity, if $A_c \approx 5b^2$. Since ρ can be approximated by $10/b^2(\tau/G)^2$ equation 5 becomes

$$D_{eff} = D_v \left[1 + \frac{10Ac}{b^2} \left(\frac{\tau}{G}\right)^2 \frac{D_c}{D_v}\right]. \tag{6}$$

Correspondingly, Eq. 4 can be written as

$$\dot{\gamma}_3 = \frac{A' D_v}{\kappa T}\left[1 + \frac{10Ac}{b^2}\left(\frac{\tau}{G}\right)^2 \frac{D_c}{D_v}\right] Gb\left(\frac{\tau}{G}\right)^n. \tag{7}$$

Diffusional Creep

At very high temperatures and for large stresses, the diffusional flow of atoms either by volume diffusion or grain boundary diffusion causes viscous creep of polycrystalline materials which includes grain boundary sliding. These additive contributions to the strain rate can be related by (Ashby 1976)

$$\dot{\gamma}_4 = A'' \frac{\tau \Omega}{\kappa T D^2} D_v \left[1 + \frac{\pi \Delta}{D}\left(\frac{D_{gb}}{D_v}\right)\right], \tag{8}$$

where A'' can vary from ~15 to 45, Ω is the atomic volume, D is the grain size, and Δ is the grain boundary thickness for effective diffusion. The volume diffusion-controlled term in Eq. 8 is known as *Nabarro-Herring creep*, while the grain boundary diffusion-controlled term is known as *Coble creep*.

Ashby (1976) has defined the net strain rate of a polycrystalline material subjected to a shear stress, τ, at a temperature, T, as

$$\dot{\gamma}(\text{net}) = \dot{\gamma}_4 + \text{greatest of } \{ \ldots . \ldots \dot{\gamma}_3, \text{least of } \dot{\gamma}_1 \text{ and } \dot{\gamma}_2 \}. \tag{9}$$

Consequently, with reference to Fig. 5a, b the maps are divided into fields such that, within a field, one contribution to Eq. 9 is larger than any other. Field boundaries therefore define a set of values τ/G and T/T_m at which dominant deformation mechanisms occur. Comparing Fig. 5a, b for Mar-M200 Ni-base superalloy illustrates that for a 100 um grain size, the dominant mechanism is Coble creep with strain rates at the typical operating conditions (shaded) in the range 10^{-8} to 10^{-10} s^{-1}, while for the much larger grain size alloy (Fig. 5b), Coble creep still dominates, but the strain rates at the operating conditions are considerably less. Deformation mechanism maps have been expanded to a large number of metals, alloys, and other materials, and the reader is referred to the more extensive compilations of Frost and Ashby (1982).

While Fig. 5a, b illustrates the utility of large grain size Ni-base superalloy (Mar-M200) for low-temperature turbine applications, Fig. 5c shows the advantages of columnar grain and single-crystal Mar-M200 (Hf) in contrast. Referring to Fig. 1c, it can be observed in Fig. 5c that there is a wide scatter (at least an order of magnitude) for creep-rupture life in the transverse columnar-grain orientations for the Mar-M200 (Hf) in contrast to <100> single-crystal longitudinal (axial) orientations over the range of roughly 675–760 °C (>0.5 T_m). This is due primarily to uncontrolled grain boundary orientations since it has been typically known that misorientations greater than about 10° result in the rapid loss of creep strength

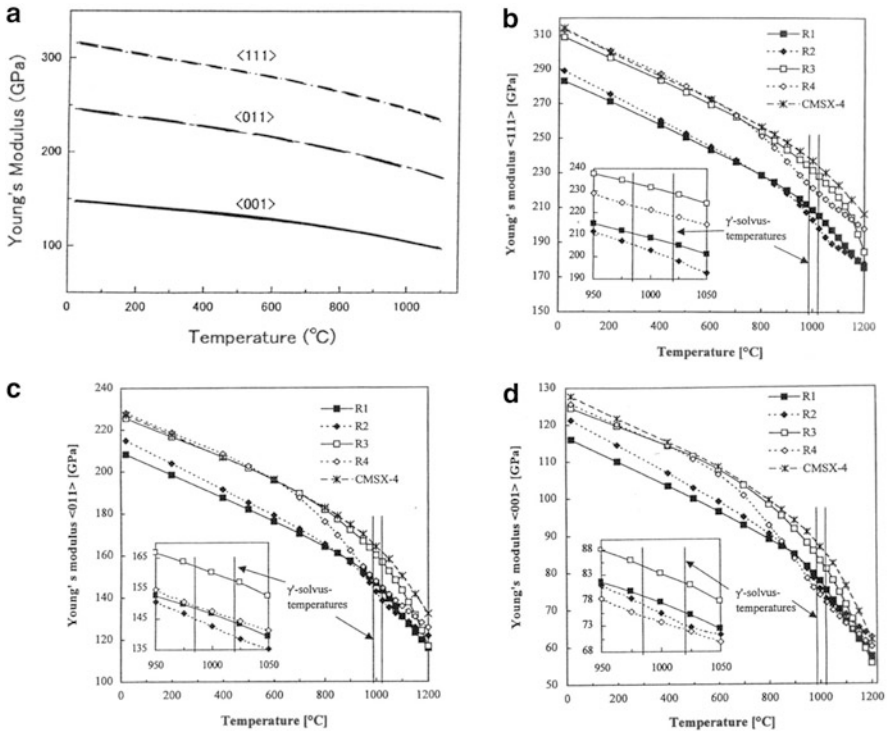


Fig. 6 Young's modulus versus temperature for Ni-base superalloy single-crystal orientations. (**a**) TMS-75 alloy (National Institute for Materials Science, Tsukuba Science City, Japan, 2004). (**b–d**) show data for experimental Ni-base alloy compared to commercial CMSX-4 alloy (Table 1). Alloy compositions are as follows: Rl-Ni-9.6Al-0.9Mo; R2-Ni-6.6Al-1.9Mo; R3-Ni-6Al-15.1Mo; R4-Ni-2.8Al-22.2Mo (balance Ni in weight percent) (Adapted from Fahrmann et al. (1999))

because of a roughly order of magnitude increase in the grain boundary diffusivity (D_{gb} in Eq. 8) in contrast to that for transgranularity (Shewman 1963).

Figure 5c also points up the creep anisotropy characteristic of single-crystal orientations, which, as implicit in ▶ Table 2 (chapter "Examples of Tensor Properties Using Matrix Fundamentals (A Physical Property)"), are also associated with elastic modulus anisotropy, especially for Ni. This is also observed for essentially all Ni-base superalloys as represented in the experimentally measured elastic moduli for Ni-base superalloys shown in Fig. 6, where the TMS-75 alloy in Fig. 6a exhibits a solvus temperature around 1,250 °C in contrast to ~1,075 °C for the alloys represented in Fig. 6b – c, where the solvus temperature represents the directional coarsening of γ′ precipitates forming a so-called rafting structure under creep conditions. The comparative creep conditions or creep curves for Ni-base superalloys, including those represented in Fig. 6, are illustrated in Fig. 7 which emphasizes the single-crystal orientation anisotropies on creep behavior and the critical role played by the initial γ′ precipitate size. It can also be observed that the

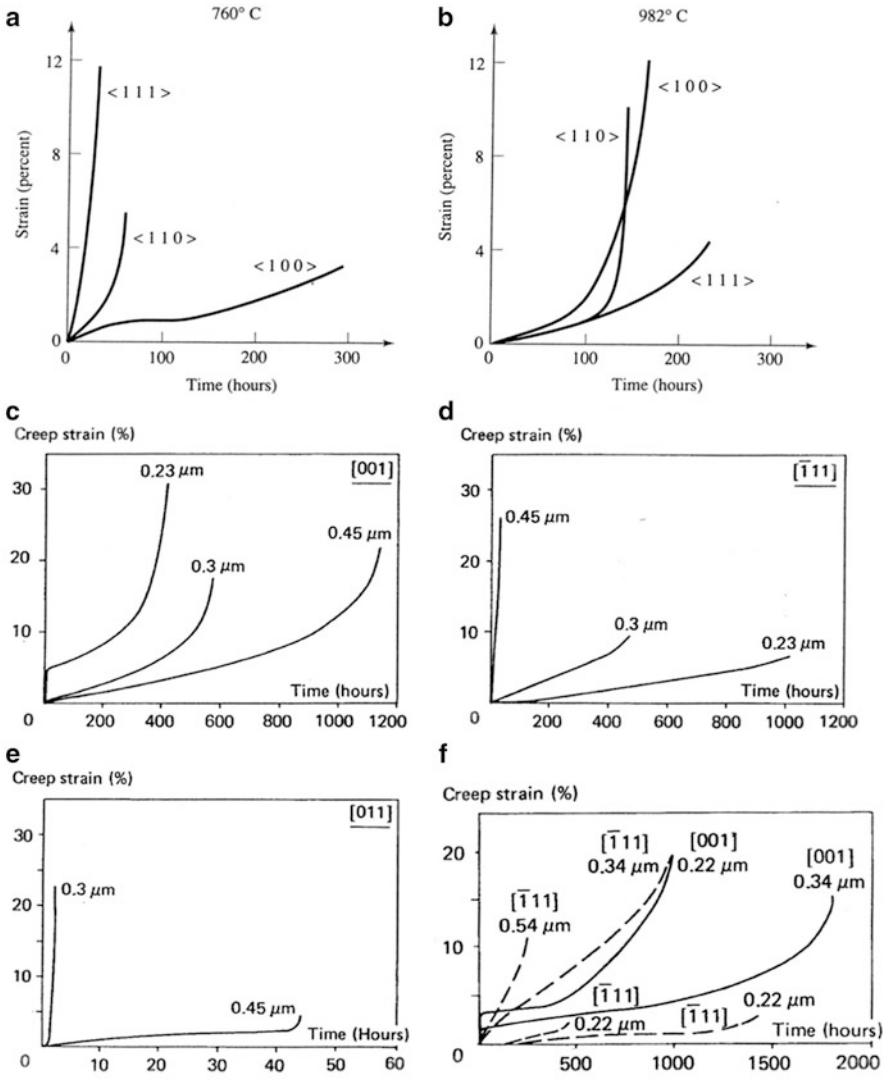


Fig. 7 Effects of orientation and γ′-phase size on creep response for single-crystal Ni-base superalloys. (**a**) and (**b**) Variation of creep response for alloy PWA1480 (Mar-M200 equivalent) at 260 °C and 982 °C, respectively. (**c**–**e**) compare CMSX-2 (Table 1) alloy creep behavior for different γ′-phase sizes, 760 °C and 750 MPa stress. (**f**) Effect of γ′-phase size on the creep behavior at 760 °C and 750 MPa stress for CMSX-4 single crystals. (**c**–**f**) are adapted from Caron et al. (1988)

creep-rupture life at 982 °C in Fig. 7b for <110> oriented and improved Mar-M200 alloy (PWA 1480) in contrast to the <110> oriented single-crystal Mar-M200 (Hf) alloy in Fig. 5c (at less than that temperature) is at least two orders of magnitude. Furthermore, the reversal in γ′ size effects shown on comparing

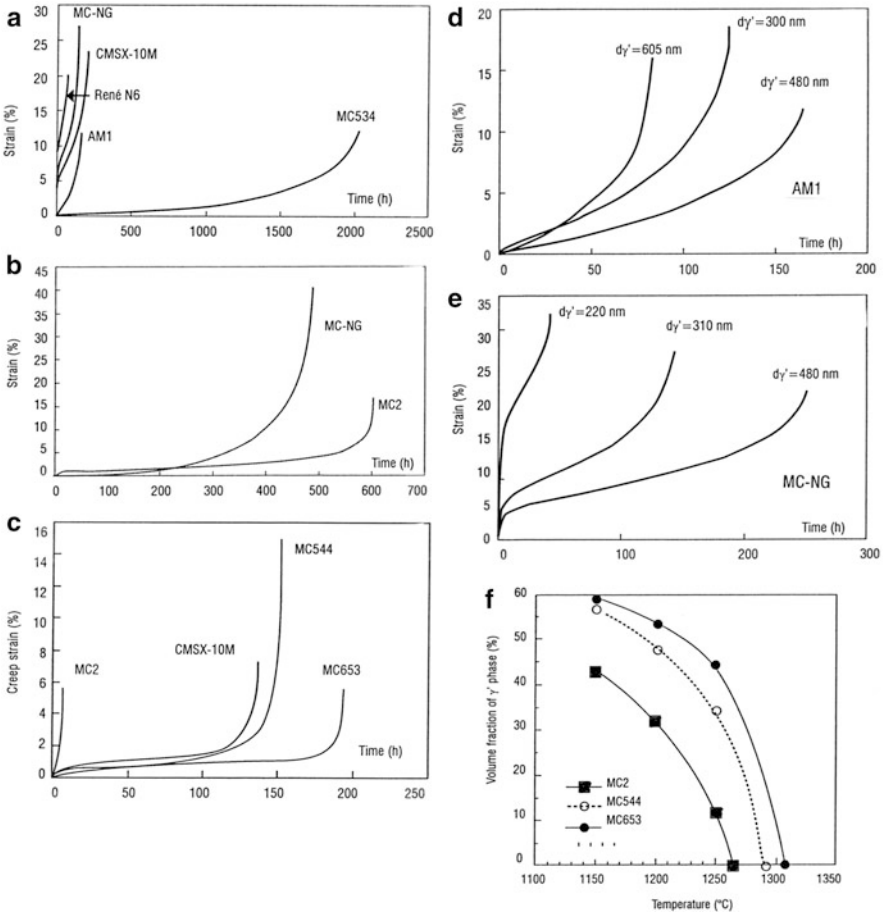


Fig. 8 Creep curves for <001> single-crystal Ni-base superalloy turbine blades. (**a**) 760 °C and 840 MPa stress for various superalloys. (**b**) 1,050 °C and 150 MPa stress for two alloys. (**c**) 1,150 °C and 100 MPa stress comparisons. (**d** and **e**) show somewhat anomalous creep behavior with various γ′-phase precipitate sizes (for (**b**)). (f) γ′-phase volume fraction changes with temperature in three single-crystal superalloys (Table 1) (Adapted from Caron and Lavinge (2011))

Fig. 7c, d, and f illustrates the complexity involved in optimizing and improving high-temperature, single-crystal, Ni-base superalloys especially for large turbine blade applications in industrial gas turbines. We emphasize this issue by perusing the comparative creep behavior for various Ni-base superalloy compositions in Fig. 8a – c (Table 1) as well as the reversal in γ′ precipitate size effects shown for two alloy compositions in Fig. 8d, e along with the changes in γ′ volume fraction in three <001> single-crystal alloys (in Fig. 8c) with temperature (having solvus temperatures from ∼1,260 °C to ∼1,310 °C).

γ and γ′-Phase Creep Deformation Fundamentals

So if we take a progressive perusal of the considerable and complex phenomena summarized in Figs. 7 and 8 (but only briefly in the context of more than a half century of Ni-base alloy and component development), it becomes clear that <001> single-crystal orientations containing varying sizes and volume fractions of the γ′-phase begin to establish a design strategy which continues to influence turbine component development, especially in the context of intermediate and high-temperature creep behavior.

Figure 9a shows a typical transmission electron microscope image corresponding to a magnified section for the columnar-grain structure in Fig. 1c as well as the single-crystal structure shown in Fig. 1d. The cuboidal γ′(Ni_3Al) precipitates range in size from ~0.1 um (100 nm) for the smallest dimension to ~0.5 um (500 nm) for the largest. Estimates of the γ′ volume fraction range from ~40 % to 55 %, and the interparticle γ-matrix channel width varies from ~25 to 75 nm. These features, γ′ size, γ − γ′ (or γ/γ′) channel size (c), γ′ volume fraction, and the γ − γ′ misfit (%), have become the principal Ni-base superalloy design features. The addition of elements which are partitioned to either γ or γ′ can either strengthen (or harden) the respective phase, increase its lattice parameter (thereby altering the misfit (► Eq. 16 in chapter "Planar Defects: Crystal Interfaces") and the corresponding coherency strain), or selectively alter the stacking-fault free energy and the corresponding dislocation reactions. Furthermore, the stability of the γ′ phase, characterized by the solvus temperature, is also an important high-temperature creep.

The γ′-phase precipitation and growth kinetics are very sensitive to the rate of cooling through the solvus temperature. Rapid cooling rates promote γ′ precipitates ~200–500 nm similar to ► Fig. 18a of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements," while slower rates promote a bimodal distribution of γ′ precipitate sizes with the smaller range having sizes <50 nm and the larger range having sizes >500 nm. The precipitates are spherical for misfit between ~0 and 0.2 %, while between 0.2 % and 0.3 %, particles evolve from spherical to rounded cuboids to well-defined cuboids for misfit > 0.3 % (Oblak and Kear 1972; Caron and Lavigne 2011; Hornbogen and Ruth 1967). This can be manipulated by elemental additions and compositions (Table 1) and is particularly notable on considering Fig. 8a where alloy MC-NG (Table 1) has a misfit of only 0.05 % in contrast to MC534 with a misfit of 1.4 % as a consequence of the higher Mo level in MC538 which partitions preferentially to the γ-matrix, significantly increasing the lattice parameter (Caron and Lavigne 2011). In addition, and as illustrated in this example, the misfit is negative. As we shall see later, a negative misfit creates rafts or rods (or γ) in a γ′ matrix perpendicular to the tensile axis in <001> single crystals in particular, in high-temperature creep.

Figure 10a illustrates the optimization of creep-rupture life for the TMS-75 alloy represented in ► Fig. 15a of chapter "Summarizing Atom and Ion

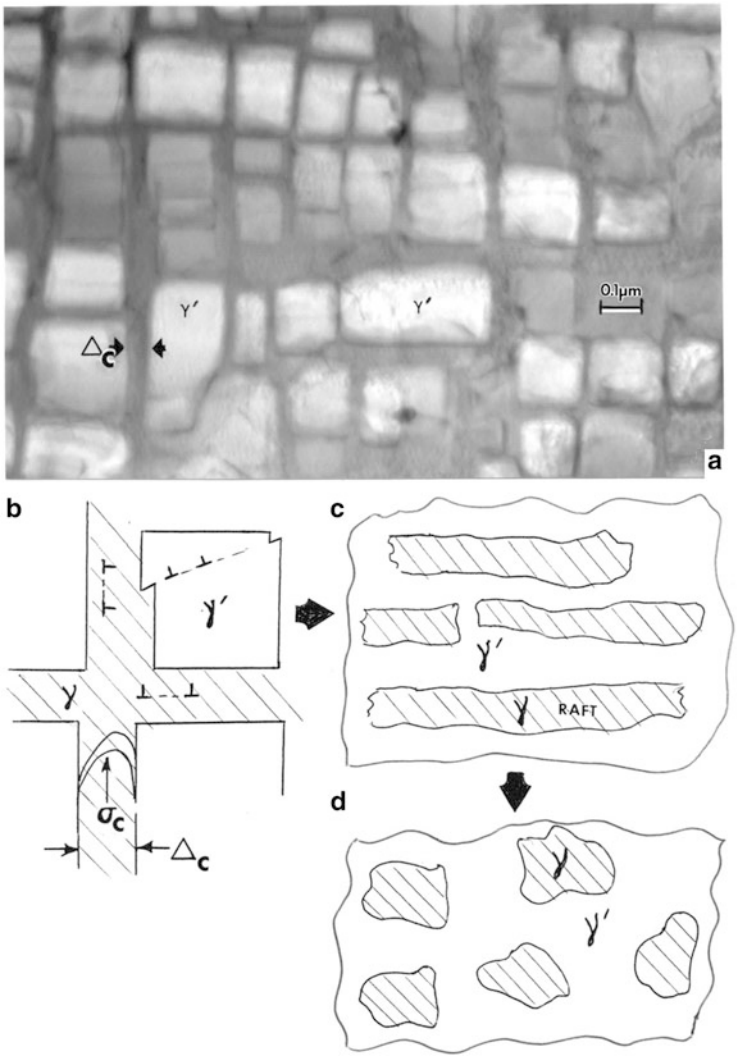


Fig. 9 γ′-phase precipitation phenomena in Ni-base superalloys. (**a**) TEM image of cuboidal γ′ precipitates in fcc γ-matrix. γ-matrix channel spacing is denoted Δ_c. (**b**) Schematic of γ − γ′ regimes and dislocation phenomena: Orowan bowing at σ_c and Δ_c, γ channel width and γ′ particle shearing. (**c** and **d**) shows coalesced γ rafts and islands in γ′ at high-temperature. Figure 9b illustrates some of these γ′ precipitate-related phenomena schematically, including dislocation glide features which are unique to creep behavior as it relates to the fcc γ-matrix, and the coherent fcc γ′ precipitate phase coincident with the matrix {001} planes: γ {001}ll γ′{001}

Structure: The Periodic Table of the Elements" for γ′-phase precipitate volume fraction at two creep stress levels and corresponding temperatures. In contrast, Fig. 10b shows that temperature dependence for the yield stress for different creep strains, and especially at higher strains, increases with temperature up to roughly

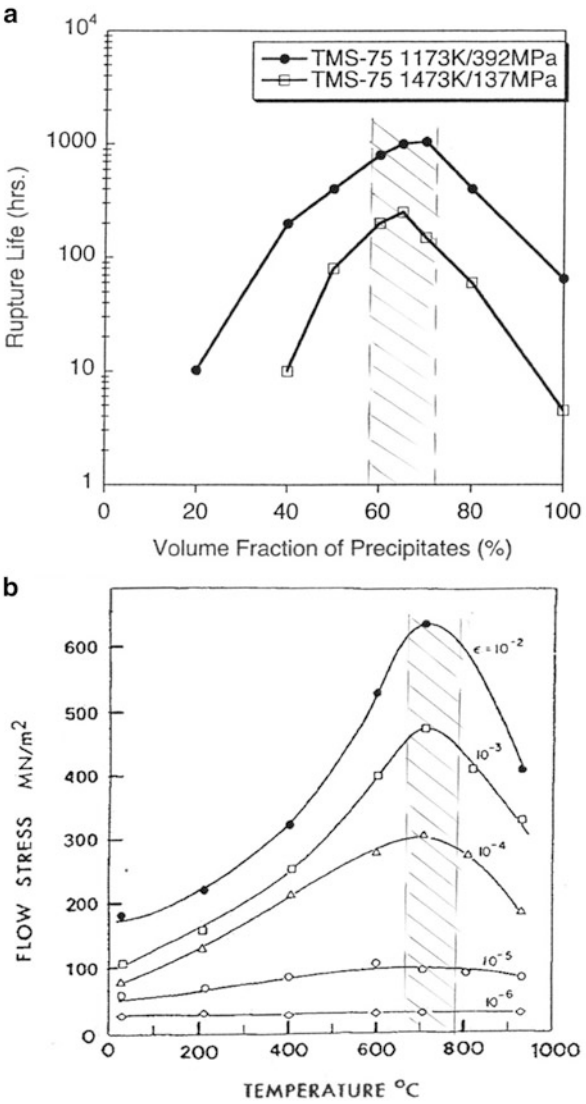


Fig. 10 γ'-phase precipitate control in creep behavior of Ni-base superalloy single-crystal turbine blades. (**a**) Creep-rupture variation of TMS-75 alloy (Table 1) versus γ'-phase volume fraction for conditions shown (Adapted from Murakumo et al. (2004)). (**b**) Yield stress versus temperature curves for γ' (Ni_3 Al) phase at creep strains shown (Adapted from Kear and Oblak (1974))

760 °C and then declines for the γ' (Ni_3Al) phase. This occurs because as dislocations enter the γ-phase, some create sessile locking arrays (Oblak and Kear 1972).

Finally, it should be recognized that like the elastic moduli, flows stress is also highly anisotropic in single crystals. This occurs because dislocations glide under the effect of a resolved shear stress acting in the slip plane, and this stress component must be determined from the normal stress acting along the crystal axis, parallel to the crystal orientation. For fcc single crystals, for example, there are 12 slip systems of the type {111} <110> and correspondingly 12 slip systems of the type {111} <112> (see ▶ Fig. 24 in chapter "Line Defects: Dislocations in

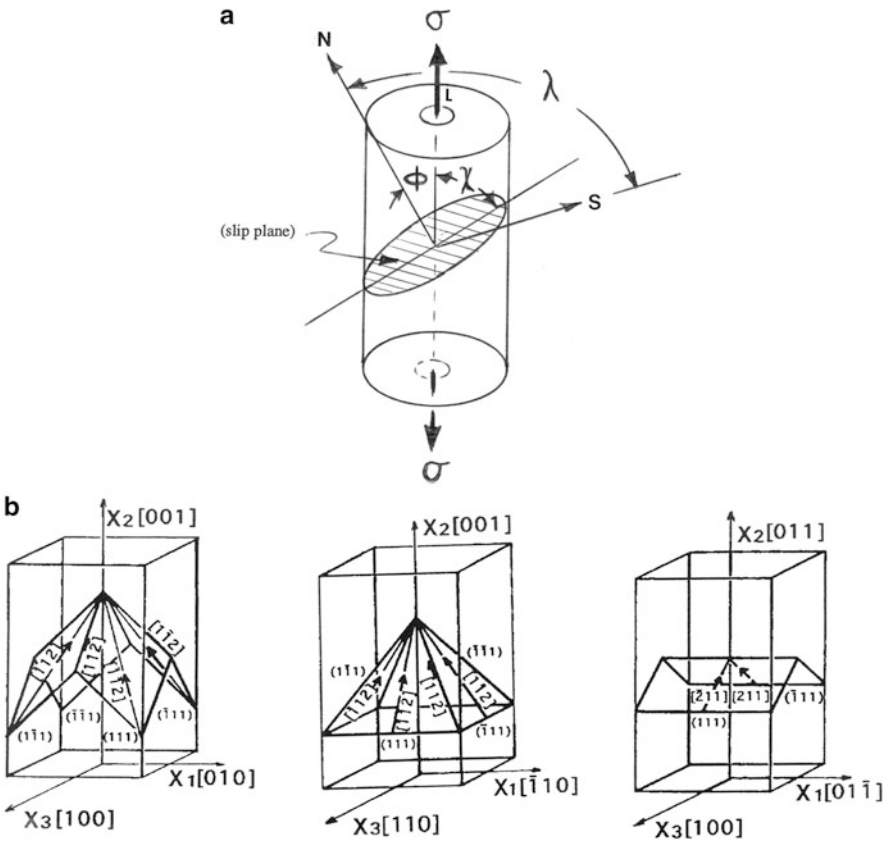


Fig. 11 Relationship between axial stress and slip plane and slip direction in a single crystal. (**a**) Normal stress (σ)-flow stress (τ) geometry. N is the normal to the slip plane. L is the axial (crystal) [hkl] direction, and s is the slip direction <hkl>. (**b**) Examples of the arrangement of {111} <112> fcc slip systems in crystal specimens (Adapted from Kakehi (2004))

Crystalline Materials"). Consequently, as shown schematically in Fig. 11a, the shear stress acting in the slip plane will be given by

$$\tau = \sigma \cos\phi \cos\lambda \tag{10}$$

where σ is the normal stress. One can also write

$$\tau_L = \sigma \sin\chi \cos\lambda = \cos\phi \cos\lambda = M\sigma \tag{11}$$

where M is the *Schmid factor*. This emphasizes the fact that metal flows plastically when the resolved shear stress acting in the slip plane and along the slip direction reaches a critical value C_c (*Schmid's law*). For low-temperature creep, {111} <112> slip systems operate in the γ-matrix, while at high temperature the {111}

<110> systems dominate. Figure 11b illustrates the arrangement of {111} <112> slip system for two specific crystal orientations. In the [001] orientation, there are 4 {111} <112> slip systems with a Schmid factor of M = 0.47. Correspondingly, there are 8 {111} <110> systems with M = 0.41. For [011], there are only two {111} <112> slip systems oriented with respect to the axial [011] systems. In the case of the $\left[\overline{1}11\right]$ oriented crystals in Fig. 7d, f the {111} <110> Schmid factor is 0.275 in contrast to 0.41 for [001], partly responsible for the apparent reversal in creep-rupture behavior for large γ' particles (0.45 um) in [001] single crystals in contrast to smaller γ' particles (0.23 um) in $\left[\overline{1}11\right]$ single crystals.

However, the most significant feature of γ'-phase optimization in Ni-base [001] single-crystal superalloys is vested in dislocation behavior in the γ-matrix versus the γ'-phase at intermediate and high-temperature creep. To examine this phenomenon, we return to Fig. 9. It might be noted that while Fig. 9a is tilted in the TEM to eliminate the $\gamma - \gamma'$ coherency strain image features, ▶ Fig. 6b in chapter "Volume Defects: 3D Imperfections in Crystals" is appropriately tilted to reveal them as δ-fringe patterns discussed in detail by Oblak and Kear (1972) and others (Gevers et al. 1964). ▶ Figure 6a of chapter "Volume Defects: 3D Imperfections in Crystals," in contrast, also shows misfit dislocation structures, in the γ/γ' interface or γ channel region as well. As implicit in Fig. 10b, the yield stress associated with the γ' increases up to the intermediate creep temperature range (~700–800 °C). Dislocations in the γ channels (Fig. 6b) bypass the γ' precipitates, and dislocation glide is controlled by Orowan bowing in the channels according to (Pollock and Tin 2006)

$$\sigma_c = \sqrt{\frac{2}{3}}(Gb/\Delta_c M), \tag{12}$$

where as noted, Δ_c is the channel size (width) and M is the Schmid factor (Eq. 11). Pollock and Tin (2006) have shown that typical values of these materials properties at 850 °C are G = 48.2 GPa, b = 0.254 nm, Δ_c = 60 nm, and M = 0.41. Correspondingly, σ_c in Eq. 13 is 408 MPa, or for small channel widths corresponding to those shown in Fig. 9a where $\Delta_c \cong 25$ um, this critical stress would rise to ~980 MPa at 850 °C for {111} <110> slip. However, from Fig. 10b, this far exceeds the γ' shear stress, and under these circumstances, dislocations would heterogeneously shear the γ'. B.H. Kear and colleagues at Pratt-Whitney in the 1960s and early 1970s very elegantly identified the dislocation reactions associated with these creep-related glide features. Of course for a low stacking-fault free energy fcc γ-matrix, we could expect

$$\frac{a}{2}<100> \rightarrow \frac{a}{6}<112> + \frac{a}{6}<112> \tag{13}$$

type reactions. But these homogeneous $\{111\}\frac{a}{2}<110>$ slip features were also observed to involve pairs of $\frac{a}{2}<112>$ dislocations coupled by an antiphase boundary (APB) forming (Oblak and Kear 1972; Kear and Oblak 1974)

$$\frac{a}{2}<110>+\frac{a}{2}<110>\rightarrow a<110>\rightarrow\frac{a}{3}<211>+\frac{a}{3}<121>, \qquad (14)$$

where the latter two dislocations form a pair of superlattice partials loosely coupled by intrinsic or extrinsic stacking faults. So while $\frac{a}{6}<112>$ partial dislocations involving $\{111\}<112>$ slip cannot easily shear the γ' precipitate block without creating a high-energy fault, the $\frac{a}{3}<112>$ superdislocations can easily enter the γ'-phase, thereby increasing the creep rate. As noted above, this occurs with decreasing channel width, Δ_c (Fig. 9b). However, the strength of γ' can also be changed by Ti and Ta additions to the alloy compositions. The development of misfit dislocation networks and related structures within the $\gamma-\gamma'$ channels as shown in ▶ Fig. 6a (chapter "Volume Defects: 3D Imperfections in Crystals") can also contribute to early creep stage development and this combined with intersecting $\{111\}<112>$ slip work hardens the channels during steady-state creep.

As creep progresses into the high-temperature regime (>900 °C) where vacancy diffusion and dislocation glide and climb processes progress, the γ'-phase begins to directionally coalesce leading to so-called rafted $\gamma-\gamma'$ microstructures normal to $<001>$ tensile stress and assumes the appearance shown in Fig. 12b in comparison with the normal, cuboidal γ'-phase shown in Fig. 12a. The rafting shown in Fig. 12b, and illustrated schematically in creep evolution in ▶ Fig. 18b of chapter "Summarizing Atom and Ion Structure: The Periodic Table of the Elements," creates an inverted microstructure relative to the original γ'-phase (Fig. 12a) which is surrounded by γ while the rafts are γ islands surrounded by the eutectic γ'-phase. At even higher temperatures, these rafts agglomerate into small islands of γ in the γ'-phase, and creep rupture occurs shortly thereafter. Consequently, tertiary creep is dominated by the topological inversion of $\gamma-\gamma'\rightarrow\gamma'-\gamma$ and dislocation climb or vacancy diffusion illustrated in Fig. 13.

Advanced Concepts for Directional Solidification

Asillustrated in Fig. 1b–d, gas-turbine blades employing directional solidification have been produced since the 1950s, more than half a century. While single-crystal blades are relatively easy to produce for small airfoils, columnar-grained blades are much cheaper and require less stringent thermo-kinetic control and are preferred for large blade designs such as that illustrated in Fig. 14 for a General Electric GTD-111 (Table 1) Ni-base superalloy employed in the first stage of high-power gas turbines (Donachie and Donachie 2002). In these directionally solidified airfoils, the preferred orientation is fcc [100] because this is the lowest energy zone axis. Variations in contrast occur in the longitudinal plane view because other orientations were presented such as [110] (or (110)) in addition to [100] (or $<100>$). It can be observed in Fig. 14 that the columnar grains vary in width from roughly 0.5 to ~2 cm, recognizing that not all of the grains in the base section are visible. In addition, and as shown in Fig. 9a, these airfoils contain γ'-phase,

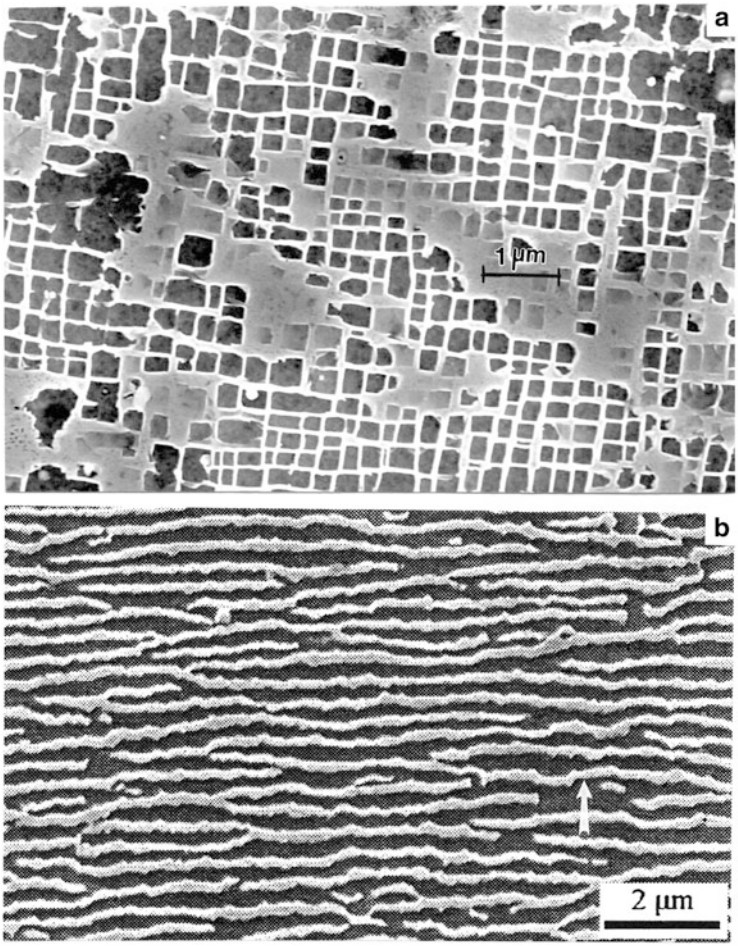


Fig. 12 SEM images of $\gamma - \gamma'$ microstructure (**a**) and microstructure inversion forming γ rafts (**b**). The microstructures are attached so that γ' is potentially attached outs leaving γ (*white*). In (**a**), γ surrounds the γ' precipitate cuboids, while in (**b**), the γ' surrounds the γ rafts. (**b**) is adapted from Polluck and Tin (2006)

cuboidal precipitates surrounded by the fcc (γ) austenitic matrix which is achieved by a complex heat treatment schedule to optimally precipitate the γ'-phase to maximize the longitudinal rupture strength while retaining good transverse blade strength and ductility.

In the case of conventional, direction solidification utilizing solidification perpendicular to a plane solid–liquid interface or solidification front, the transition from the melt or liquid phase to the solid phase, the solid phase must have a lower free energy than the liquid phase at temperatures below the melting point. Correspondingly, the liquid phase free energy must be lower above the melting point,

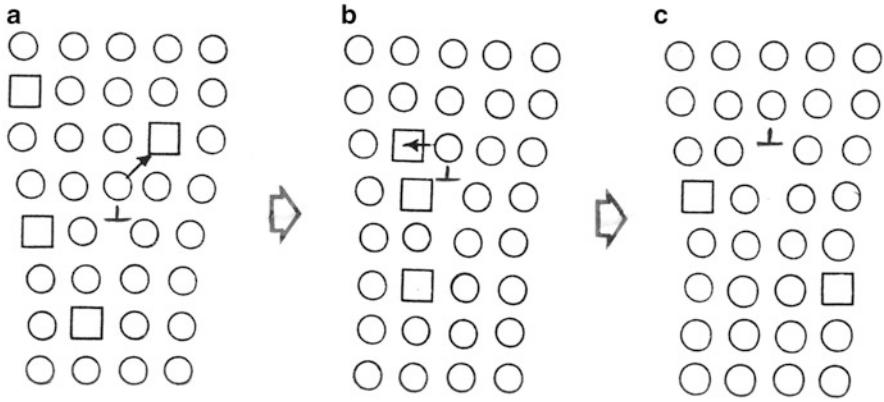


Fig. 13 Cartoon showing dislocation climb with absorption of vacancies at dislocation core in sequence A → B → C. Note progression of extra-half plane (dislocation) perpendicular to the slip or glide direction. Movement of atoms to vacancies are shown by *arrows*

Fig. 14 General Electric high-power gas-turbineairfoil composed of directional solidified GTD 111 Ni-base superalloy exhibiting large columnar-grain structure

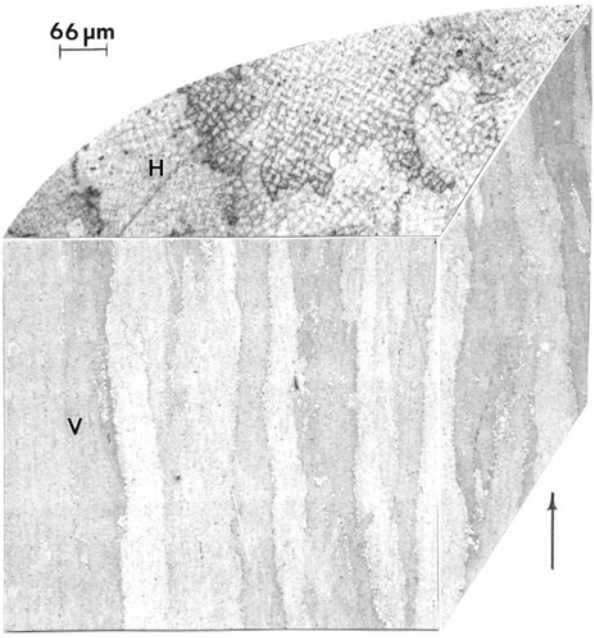


Fig. 15 3-D composition showing optical microscope view of a section from an additively manufactured (electron beam melting) Ni-base superalloy (Rene′ 142) showing columnar-grain structures directionally oriented in the build direction indicated by the *arrow*. H and V indicate horizontal and vertical reference planes

which is defined as the temperature at which the free energies for the solid and liquid phases are equal and coexist in equilibrium. During solidification, the liquid just in front of the solid–liquid interface is actually at a temperature below the liquid equilibrium temperature and correspondingly supercooled. This constitutional supercooling arises from a composition change rather than a temperature change.

Recently, a new concept in directional solidification has been developed using electron or laser beam melting of powder layers, one layer at a time. This builds a layer-by-layer monolith which involves layer-by-layer melting and cooling: an interrupted solidification front having dimensions commensurate with the layer thickness (Murr et al. 2012a, b, c). These additive manufacturing or 3-D printing/fabrication concepts will be described in more detail later, but it seems appropriate to at least highlight the prospects for building complex, monolithic airfoils using these concepts.

Figure 15 illustrates a small section extracted from a Rene′ 142 alloy (Table 1) component fabricated by electron beam melting using pre-alloyed powder. This alloy, also an evolving General Electric Corporation Ni-base superalloy similar to the GTD 111 alloy illustrated in Fig. 14, has also been used in commercial gas-turbine blade production since 1991. Figure 15 shows columnar grains similar to those in Fig. 14, but on a scale a factor of 100 smaller, while Fig. 16 illustrates the preferred orientations in the horizontal and vertical reference planes to be [100] ([200]) and [100] and [110] ([220]), respectively. The columnar-grain sizes intersecting the horizontal reference in Fig. 15 are illustrated in Fig. 17 which illustrates the grain boundaries and the grain boundary misorientations, as well as

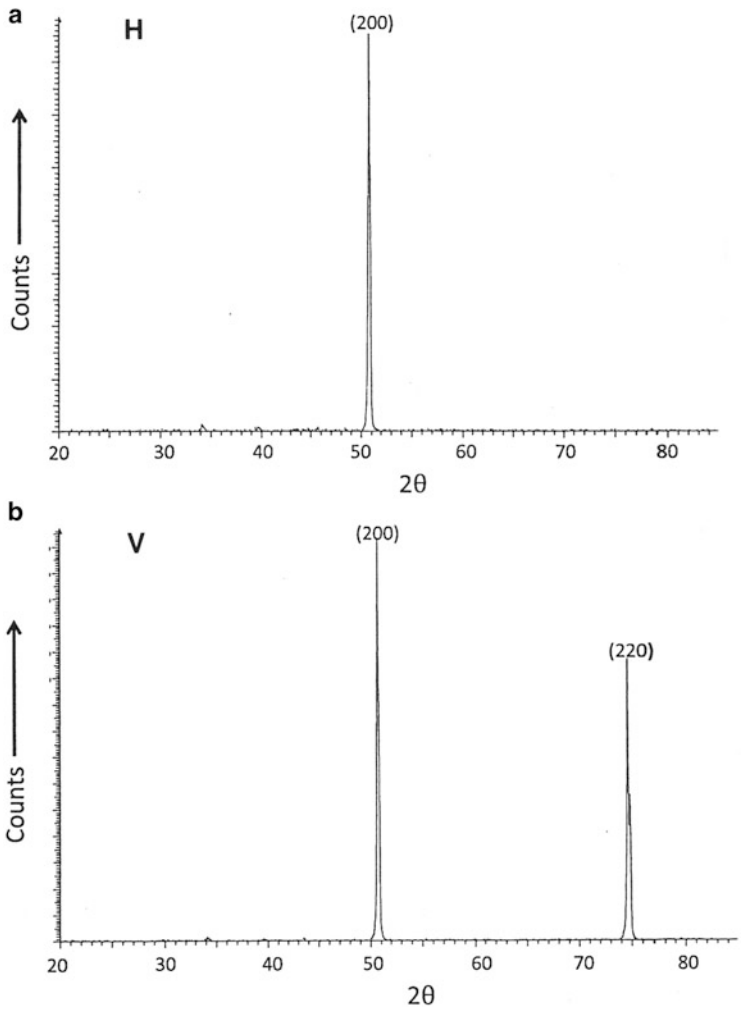


Fig. 16 XRD spectra corresponding to (**a**) the horizontal (H) and (**b**) the vertical (V) reference planes shown in Fig. 15

the regular $\gamma - \gamma'$ microstructures. HfC precipitates are also visible in the grain boundaries and the $\gamma - \gamma'$ matrix. These arise because of the carbon included in the composition shown in Table 1. Figure 18 shows a view of the $\gamma - \gamma'$ microstructure for comparison with Fig. 9a. It can be observed in this comparison that the γ'-phase precipitates are more regular and have a larger volume fraction near the optimum as illustrated in Fig. 10a. This feature poses interesting prospects for fabricating complex hollow or other intricate airfoil cooling configurations having optimal microstructures contributing to optimal creep strength behavior without the conventional heat treatment schedule (Donachie and Donachie 2002;

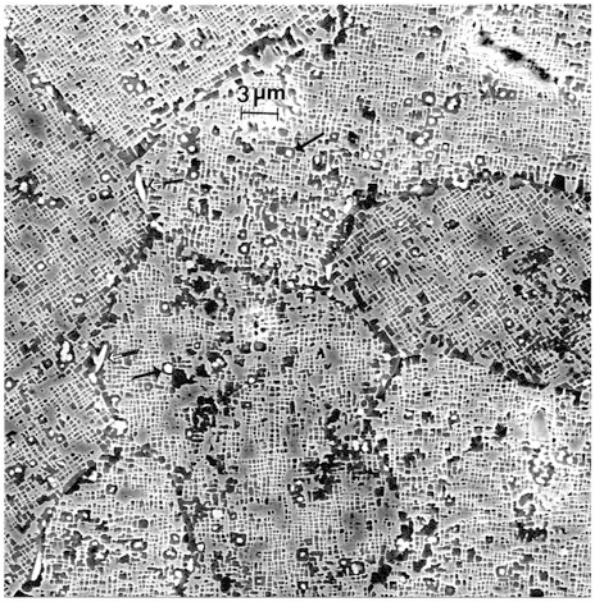


Fig. 17 Magnified horizontal (*H*) SEM section view in Fig. 14 showing γ′-phase cuboidal precipitates (selectively etched out). HfC precipitates are shown by *arrows*in the grain boundaries and matrix. Grain misorientations noted by the angles between <100> cuboidal precipitate directions vary from ~2° to 52°

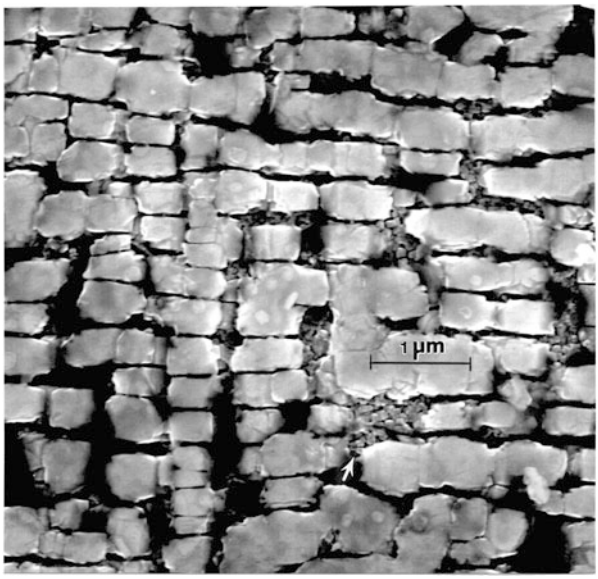


Fig. 18 SEM view of the γ′-cuboidal precipitates in a vertical section (V) corresponding to Fig. 14. The etching selectively removed the γ-matrix surrounding the precipitates, the reverse of Fig. 17. *Arrow* shows regions of very small γ′-phase precipitates

Ross and O'hara 1992). Figure 18 corresponds to the TEM images shown in ► Fig. 6 in chapter "Volume Defects: 3D Imperfections in Crystals".

The development of turbine airfoils, especially for very large turbine applications where their length can be much larger than that shown in ► Fig. 23 of chapter

"Summarizing Atom and Ion Structure: The Periodic Table of the Elements" and at even increasing operating temperatures, is an ongoing materials science and engineering challenge. As the operating temperatures approach the melting point of existing superalloy systems, innovations in airfoil cooling utilizing novel fabrication concepts such as additive manufacturing employing computer-aided design strategies may provide important extensions for existing and new alloy compositions.

References

Ashby MF (1972) A first report on deformation mechanism maps. Acta Metall 20:887–897

Ashby MF (1973) The microstructure and design alloys.In: Proceedings of 3rd international conference on strength of metals and alloys, vol 2. Cambridge University Press, Cambridge, UK, p 8

Ashby MF (1976) Chapter 11: Deformation mechanism maps applied to the creep of elements and simple inorganic compounds. In: Murr LE, Stein C (eds) Frontiers in materials science. Marcel Dekker, New York, pp 391–419

Caron P, Lavigne O (2011) Recent studies at Onera on superalloys for single crystal turbine blades. J Ahaspace Lab, Novi (3) AL03-02: 1–14

Caron P, Ohta Y, Nakagawa YG, Khan T (1988) Creep deformation anisotropy in single crystal superalloys. In: Reichman S, Duhl DM, Maurer G, Antolovich S, Lund C (eds) Superalloys 1988. TMS, Warrendale, pp 215–224

Donachie MJ, Donachie SJ (2002) Superalloys: a technical guide, 2nd edn. ASM International, Materials Park

Fahrmann M, Hermann W, Fahrmann E, Boegli A, Pollock TM (1999) Determination of matrix and principate elastic constants in ($\gamma - \gamma'$) Ni-base model allots and their relevance to rafting. Mater Sci Eng A260:212–221

Frost HJ, Ashby MF (1982) Deformation mechanism maps: the plasticity and creep of metals and ceramics. Pengamon Press, New York

Gevers R, Delavignette P, Blank H, Amelinckx S (1964) Electron microscope transmission images of coherent domain boundaries I. Dynamical theory. Phys Stat Sol (b) 4:383–410

Hornbogen E, Roth M (1967) Die verteilong kohärenter teilchen in nickellegierungen. Z Mellakunde 58:842–855

Kakehi K (2004) Influence of primary and secondary crystallographic orientations on strengths of nickel-based superalloy single crystals. Mater Trans 45(6):1824–1828

Kear BH, Oblak JM (1974) Deformation modes in γ' precipitation hardened nickel-base alloys. J de Phys 35(12): 7–35–45

Morakumo T, Kobayashi T, Koizumi Y, Hanada H (2004) Creep behavior of Ni-base single-crystal superalloys with various gamma′ volume fraction. Acta Mater 52(12): 3737–3744

Murr LE, Martinez E, Gaytan SM, Ramirez DA (2012a) Contributions of light microscopy to contemporary materials characterization: the new directional solidification. Metallug Microstruct Anal 1:45–58

Murr LE, Gaytan SM, Ramirez DA, Martinez E, Hernandez J, Amato KN, Shindo PW, Medina FR, Wicker RB (2012b) Metal fabrication by additive manufacturing using laser and electron beam melting technologies. J Mater Sci Technol 28(1):1–14

Murr LE, Martinez E, Amato KN, Gaytan SM, Hernandez J, Ramirez DA, Shindo PW, Medina F, Wicker RB (2012c) Fabrication of metal and alloy components by additive manufacturing: examples of 3D materials science. J Mater Res Technol 1(1):42–54

Oblak JM, Kear BH (1972) Analysis of microstructures in nickel-base alloys: implications for strength and alloy design. In: Thomas G, Fulrath RM, Fisher RM (eds) Electron microscopy and structure of materials. University of California Press, Berkeley, pp 566–616

Pollock TM, Tin S (2006) Nickel-based superalloys for advanced turbine engines: chemistry, microstructure, and properties. J Propulsim Power 22(2):361–374

Ross EW, O'hara KS (1992) Rene 142: a high strength, oxidation resistant DS turbine airfoil alloy. In: Antolovich SD, Stusrud RW, Mackay RA, Anton DC, Khan T, Kissinger RD, Klarstrom DL (eds) Superalloys 1992. The Minerals, Metals and Materials Society (TMS), Warrendale, pp 257–265

Shah DM, Cetel A (2000) Evaluation of PWA 1483 for large single crystal IGI blade applications. In: Pollock TN, Kissinger RP, Bowman RR et al (eds) Superalloys 2000. TMS, Warrendale

Shewman PG (1963) Difussion solids. McGraw-Hill, New York, p 173

Sims CT, Hagel C (eds) (1972) The superalloys. Wiley, New York, p. vii

Part VII

Composite Materials and Structures

Classification of Composite Materials and Structures

Contents

Abstract

Composite materials in the simplest case are composed of some phase (a second phase) included in a matrix. The properties of the matrix are altered depending upon the volume fraction of this phase, as well as its geometry and orientation relative to an applied stress.

Introduction

Look around. The world is a composite. All living systems are composites. Prominent among them in humans are bones whose porous structure reduces density (or weight) and whose matrix must be stiff and strong and able to withstand impact among other functions. Composite materials are designed to achieve multiple functions which when integrated into a system forming extended structures expands these features on a larger scale. In the simplest case composites are represented by regimes or phases (and interfaces) of varying morphologies, volume fractions, structures, and properties which compose a volume element. In contrast to a single phase represented by the same volume element or reference element, the composite element is differentiated by interphase interfaces, which increase significantly for nanocomposites.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_23

Narrative

In general, composite materials are isotropic (or quasi-isotropic) or anisotropic either by circumstance or by design. The simplest are aggregated composites such as concrete or asphalt, both having various particulate shapes and sizes in a low fraction binder or matrix. For example, the volume fraction of the cement matrix or the pitch matrix, respectively, is around 10 %. In other particle-matrix composites, such as the 2 vol.% ThO_2 in Ni shown in ▶ Fig. 7 in chapter "Strengthening by Crystal Imperfections", the matrix is the high volume fraction. Composites represented by ▶ Fig. 7 of chapter "Strengthening by Crystal Imperfections" are isotropic, and the dispersed ThO_2 phase provides strengthening as implicit in ▶ Fig. 5f (chapter "Strengthening by Crystal Imperfections"). In these general cases, which can include the 2-phase system of hard (or strong) Be particles in an Al matrix as shown in ▶ Fig. 34 (chapter "Planar Defects: Crystal Interfaces"), the strengthening of a 2-phase composite can be expressed very approximately by

$$\sigma_c = V_\alpha \sigma_\alpha + V_\beta \sigma_\beta, \tag{1}$$

where α and β and the 2 phases, V_α, V_β, and σ_α, σ_β, are the corresponding volume fractions ($V_\alpha + V_\beta = 1$) and yield strengths (or flow stress) for the respective phase. Equation 1 assumes that one phase (the dispersoid or phase dispersed in a matrix) is much harder or stronger than the other (matrix phase). This is indeed the case for ▶ Fig. 34 of chapter "Planar Defects: Crystal Interfaces" and ▶ Fig. 7 of chapter "Strengthening by Crystal Imperfections", where both the included phase size and volume fraction differ significantly as well. These features are illustrated in Fig. 1a, b.

Figure 1c, d represents two examples of discontinuous fibers in a unidirectional (aligned) and correspondingly anisotropic arrangement and a randomly oriented, isotropic (or quasi-isotropic) arrangement, respectively. Figure 1c is characteristic of the superconducting wire composites shown previously in ▶ Fig. 44 of chapter "Electromagnetic Color and Color in Materials." In contrast, Fig. 1e, f shows continuous, unidirectional fibers parallel and/or perpendicular to an applied stress axis or at a specific angle relative to these axes, respectively. Figure 1g–i represents variances of the continuous fiber composite shown in Fig. 1e, f representing alternate planes of oriented fibers or fiber weaves (Fig. 1g, h) and randon, continuous fiber regimes represented by Fig. 1i. Figure 1g–i represents so-called cross-ply systems, examples of which are represented in Fig. 2a–e. Simple examples of such cross-ply composites involve plywood, which, as wood/laminate, was first utilized by the Egyptians around 3000 B.C. Similarly, random, short fiber composites in antiquity involve horse hair in plaster and stucco and straw in clay blocks, although in these composites, the fibers were utilized more as binders rather than strengthening agents. The use of palygorskite nanofibers in Mayan paintings served as a binder as well as a host for the indigo molecules in the case of Mayan blue coloration illustrated in ▶ Fig. 9 in chapter "Examples of Materials Science and Engineering in Antiquity." In the case of strong fibers in cement, for example, asbestos fibers as shown in ▶ Fig. 2 (chapter "Examples of Materials Science and

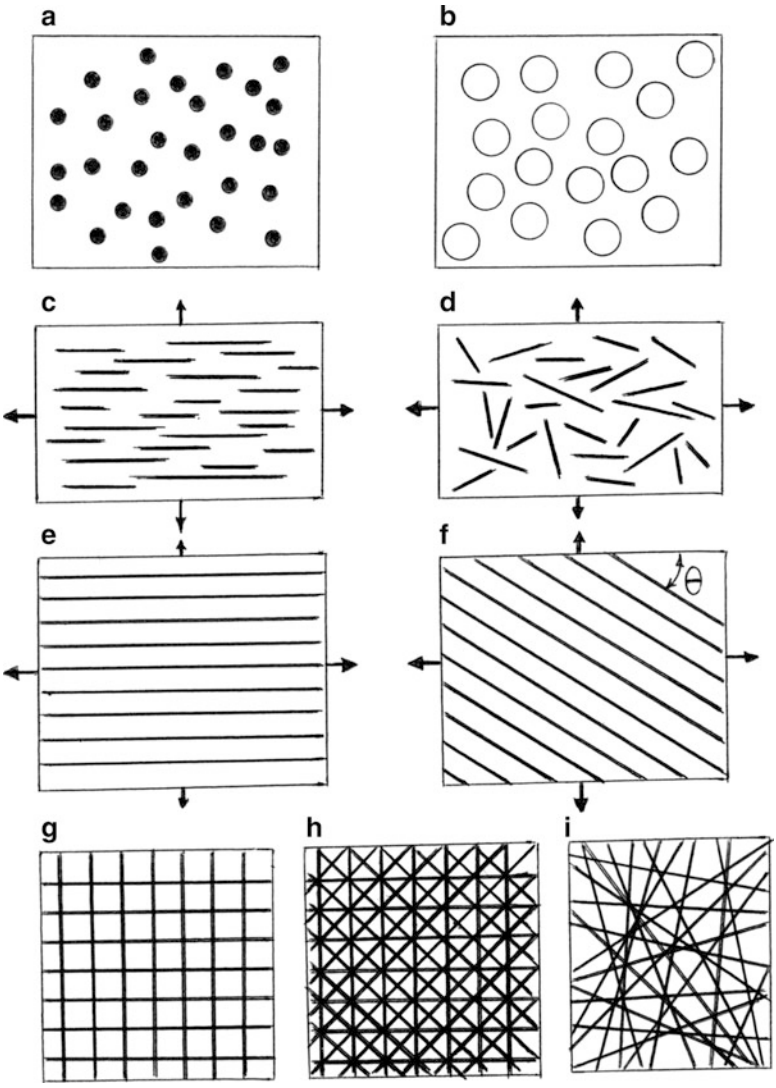


Fig. 1 Composite material classification examples. (**a**) and (**b**) show homogeneous particle distribution in a matrix. (**c**) and (**d**) show aligned (anisotropic) and random (quasi-isotropic) discontinuous fibers. (**e**) and (**f**) show aligned (unidirectional) continuous fibers: $\theta = 0°$, $\theta = 90°$; $0 < \theta < 90$. (**g**) bidirectional or cross-ply continuous fibers. (**h**) multidirectional fibers. (**i**) continuous, random

Engineering in Antiquity"), the breaking strain is much higher than the cement matrix so the fibers hold the matrix together when it cracks since the fibers can bridge the cracks. Even small fiber volume fractions can serve to eliminate catastrophic failure of weaker and more brittle matrices by restraining crack opening

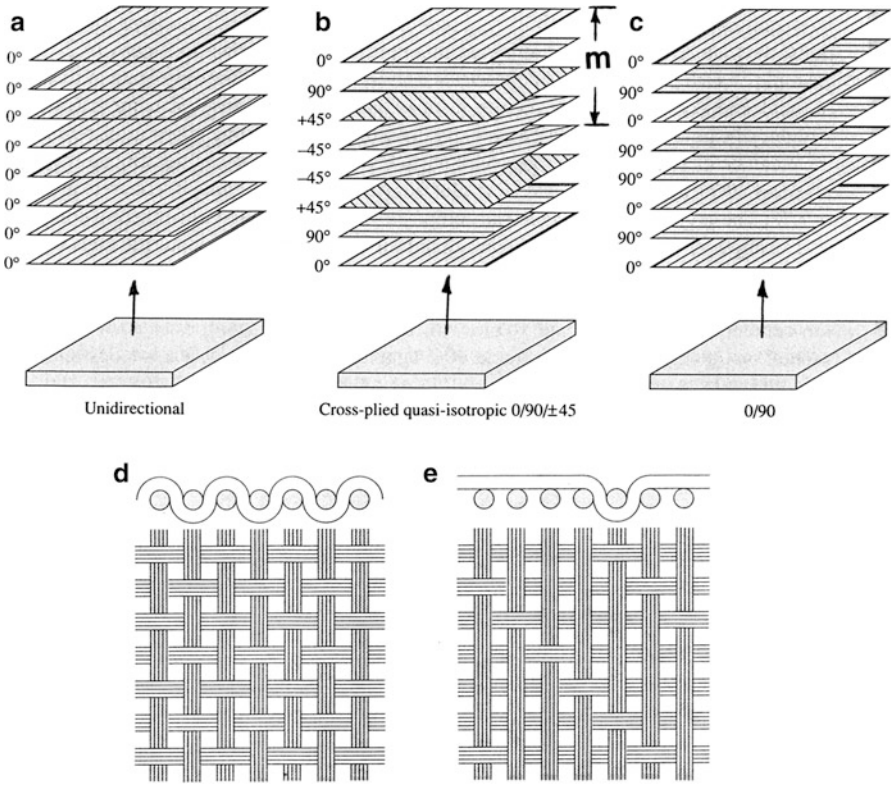


Fig. 2 Laminate fiber composite arrangements (**a**–**c**), and 2-D fabric weave architectures (**d**) and (**e**). m in (**b**) denotes continuous, multidirectional laminate (▶ Fig. 1h of chapter "Chemical Forces: Molecules"). (**d**) shows a plain weave, while (**e**) shows a 5-harness satin weave, illustrating multiplicity of fabric weave architecture (Adapted from Chawla (1987))

and propagation while providing continued or residual strength by holding the cracked matrix together (Kelly 1976).

It can be observed in Fig. 1c that for very small fibers having a fiber length/fiber diameter ratio $(l/d)_f < 10$, their alignment in a principal direction will have little consequence since they would only strengthen the matrix similar to Eq. 1 for dispersed particulate composites (Fig. 1a, b). However, a critical fiber aspect ratio, $(l_c/d)_f$, is sometimes defined as

$$(1_c/d)_f = \sigma_f'/2\,\tau_m, \tag{2}$$

where l_c is a critical fiber length, d is the fiber diameter, σ_f' is the ultimate tensile strength of the fiber material, and τ_m is the flow stress of the matrix. Generally, for fiber lengths $l > l_c$:

$$\sigma_c \cong \sigma_f^* V_f \left(1 - \frac{l_c}{2l}\right) + \sigma_m (1 - V_f), \tag{3}$$

where σ_c is the composite strength (yield stress), σ_f^* is the fiber fracture stress, V_f is the volume fraction of fiber of length l, and σ_m is the yield stress for the matrix.

For fiber length $l < l_c$:

$$\sigma_c \cong \frac{l\tau_c}{d} + V_f + \sigma_m (1 - V_f), \tag{4}$$

where τ_c is the smaller of either the matrix or the fiber flow stress and d is the fiber diameter.

For random, discontinuous fiber composites (Fig. 1d), where $l/d > 10^2$, Eq. 1 can be approximated by

$$\sigma_c \cong \sigma_m V_m + K\sigma_f V_f \tag{5}$$

or

$$E_c \cong E_m V_m + K E_f V_f, \tag{6}$$

where σ_m and σ_f are the matrix and fiber yield strengths, respectively; V_m and V_f are the corresponding volume fractions, E_m and E_f are the Young's moduli; and K is an efficiency factor:

$$K = 3/8 \text{ for a random } 2 - D \text{ composite}$$

$$K = 1/5 \text{ for a random } 3 - D \text{ composite}$$

for discontinuous fiber composites.

In treating continuous ($1 \to \infty$) fiber reinforcement of a matrix as illustrated in Fig. 1e, f, we can consider an orthotropic or transversely isotropic 3-D composite when one of its principal planes is a plane of isotropy as illustrated in Fig. 3a or a more simplified, thin, unidirectional lamina under a state of plane stress implicit in Fig. 3b. For the general orthotropic (and transversely isotropic) case in Fig. 3a, the stress–strain relations can be expressed by

$$\begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ \nu_{23} \\ \nu_{31} \\ \nu_{12} \end{bmatrix} = \begin{bmatrix} \frac{1}{E_1} & -\frac{\nu_{21}}{E_2} & \frac{\nu_{31}}{E_3} & 0 & 0 & 0 \\ -\frac{\nu_{12}}{E_1} & \frac{1}{E_2} & -\frac{\nu_{32}}{E_3} & 0 & 0 & 0 \\ -\frac{\nu_{13}}{E_1} & \frac{\nu_{23}}{E_2} & -\frac{1}{E_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{G_{23}} & 0 & 0 \\ 0 & 0 & 0 & & \frac{1}{G_{13}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{G_{12}} \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \tau_{23} \\ \tau_{31} \\ \tau_{12} \end{bmatrix} \tag{7}$$

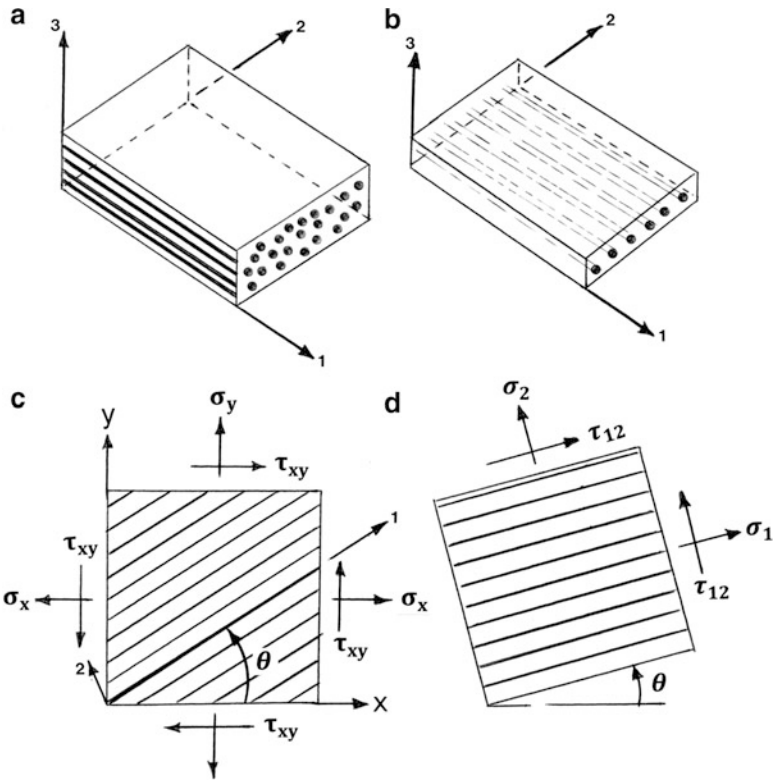


Fig. 3 (**a**) shows an orthotropic composite (continuous fibers in a matrix) with transverse isotropy, while (**b**) shows a corresponding unidirectional lamina with same principal coordinate axes, *1*, *2*, and *3*. (**c**) and (**d**) show stress components in unidirectional lamina referred to loading (*x*, *y*) and material (*1*, *2*) axes as in Fig. 1f

where γ_4, γ_5, and γ_6 are the shear strains; τ_4, τ_5, and τ_6 are the shear (or flow) stresses; and the v_{ij} represents Poisson's ratio where in general

$$\frac{v_{ji}}{E_i} = \frac{v_{ji}}{E_j} \text{or} \frac{v_{ij}}{v_{ij}} = \frac{E_i}{E_j} (i, j = 1, 2, 3) \tag{8}$$

Consequently in Eq. 7:

$$\left.\begin{aligned} \frac{v_{12}}{E_1} &= \frac{v_{21}}{E_2} \\ \frac{v_{13}}{E_1} &= \frac{v_{31}}{E_2} \\ \frac{v_{23}}{E_2} &= \frac{v_{31}}{E_3} \end{aligned}\right\} \tag{9}$$

For the orthotropic, transversely isotropic case in Fig. 3a, the plane 2–3 is the plane of isotropy, and

$$\left.\begin{array}{l} E_2 = E_3 \\ G_{12} = G_{13} \\ v_{12} = \gamma_{13} \\ v_{21} = \gamma_{31} \\ v_{23} = \gamma_{32} \end{array}\right\} \tag{10}$$

From Eq. 10, Eq. 7 can be expressed as

$$\left.\begin{array}{l} \in_1 = \left({}^{1}/_{E_1}\right)[\sigma_1 - v_{12}(\sigma_2 + \sigma_3)] \\ \in_2 = \left({}^{1}/_{E_2}\right)[\sigma_2 - (v_{21}\sigma_1 + v_{23}\sigma_2)] \\ \in_3 = \left({}^{1}/_{E_2}\right)[\sigma_3 - (v_{21}\sigma_1 + v_{23}\sigma_2)] \end{array}\right\} \tag{11}$$

In addition,

$$G_{23} = E_2/2/1(+v_{23}) = \tau_{23}/v_{23} \tag{12}$$

which reduces the number of independent elastic constants in Eq. 11 to five. If the unidirectional composite were thin in the three direction in Fig. 3b for the plane stress lamina, two-dimensional orthotropic equations replace Eq. 11 in the form

$$\left.\begin{array}{l} \in_1 = \left({}^{1}/_{E_1}\right)(\sigma_1 - v_{12}\sigma_2) \\ \in_2 = \left({}^{1}/_{E_2}\right)[\sigma_2 - v_{21}\sigma_1] \\ \gamma_{12} = {}^{\tau_{12}}/_{G_{12}} E_2 v_{12} = E_1 v_{21} \end{array}\right\} \tag{13}$$

and Eq. 13 utilizes only four independent material constants: E_1, E_2, v_{12}, and G_{12}.

For the bidirectional lamina shown in Fig. 1g, Eq. 13 becomes

$$\left.\begin{array}{l} \in_1 = \left({}^{1}/_{E_1}\right)(\sigma_1 - v_{12}\sigma_2) \\ \in_2 = \left({}^{1}/_{E_2}\right)[\sigma_2 - v_{21}\sigma_1] \\ \in_1 = E_2, v_{12} = \tau_{12}/G_{12} \end{array}\right\} \tag{14}$$

which involve only three independent elastic constants: E_1, γ_{12}, and G_{12}. This is further simplified for the continuous, multidirectional lamina represented in Figs. 1h and 2b representing a cross-plied, quasi-isotropic lay-up composite:

$$E_1 = E_2, v_{21} = v_{12}, G_{12} = E_1/2(1 + v_{12}) \tag{15}$$

which has only two independent elastic constants.

We can observe in Fig. 3c, d, corresponding to Fig. 1f, that lamina principal axes (1, 2) do not coincide with the loading or reference axes (x, y), and the principal axes (1, 2) can be expressed in terms of those referred to the loading axes (x, y) as follows:

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \tau_{12} \end{bmatrix} = [\mathrm{T}] \begin{bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{bmatrix} \tag{16}$$

and

$$\begin{bmatrix} \mathrm{E}_1 \\ \mathrm{E}_2 \\ \nu_{12}/2 \end{bmatrix} = [\mathrm{T}] \begin{bmatrix} \mathrm{E}_x \\ \mathrm{E}_y \\ \gamma_{xy} \end{bmatrix} \tag{17}$$

where the transformation matrix [T] is given by

$$[\mathrm{T}] = \begin{bmatrix} m^2 & n^2 & 2mn \\ n^2 & m^2 & -2mn \\ -mn & mn & m^2 - n^2 \end{bmatrix} \tag{18}$$

and m = cos θ, and n = sin θ in Fig. 3c and (d). From this notation one can write

$$\left.\begin{aligned} \frac{1}{E_x} &= \frac{m^2}{E_1}\left(m^2 - n^2\nu_{12}\right) + \frac{n^2}{E_2}\left(n^2 - m^2\nu_{21}\right) + \frac{m^2n^2}{G_{12}} \\ \frac{1}{E_y} &= \frac{n^2}{E_1}\left(n^2 - m^2\nu_{12}\right) + \frac{m^2}{E_2}\left(m^2n^2\nu_{21}\right) + \frac{m^2n^2}{G_{12}} \\ \frac{1}{G_{xy}} &= \frac{4m^2n^2}{E_1}\left(1 + \nu_{12}\right) + \frac{4m^2n^2}{E_2}\left(1 + \nu_{21}\right) + \frac{\left(m^2 - n^2\right)^2}{G_{12}} \end{aligned}\right\} \tag{19}$$

For basic lamina properties E_1, E_2, G_{12}, and ν_{12} and the Young's modulus, E_x, at an angle θ = 45° with the fiber direction (Fig. 3c), one can obtain from Eq. 19:

$$\left(\frac{1}{\mathrm{E_x}}\right)_{\theta=45^\circ} = \frac{1 - \nu_{12}}{4E_1} + \frac{1 - \nu_{21}}{4E_2} + \frac{1}{4G_{12}} \tag{20}$$

which can be simplified for the case of a high stiffness composite where

$$\mathrm{E}_1 >> \mathrm{E}_2 \text{ and } \nu_{21} << 1;$$

and Eq. 20 becomes

$$\mathrm{E_x}(\mathrm{q} = 45^\circ) = 4\,\mathrm{G}_{12}\mathrm{E}_2/(\mathrm{G}_{12} + \mathrm{E}_2) \tag{21}$$

indicating that the Young's modulus at 45° with the fiber direction (or axis) is a matrix-dominated property since it depends on E_2 and $\mathrm{G}_{12,}$ which reflect matrix properties.

We can now observe in Fig. 3b that properties related to the composite lamina, or for that matter the orthotropic composite in Fig. 3a in the longitudinal (or fiber) direction, can be expressed by

$$E_l = E_c = V_f E_f + V_m E_m \tag{22}$$

which is the classic law of mixtures formula or the Voigt (1889) *model* [217] uniform strain (isostrain) in a longitudinal composite:

$$\in_c = \in_m = \in_f \tag{23}$$

or

$$E_c = E_1 V_1 + E_2 V_2 + E_3 V_3 + \ldots E_i V_i \tag{24}$$

where E_i represents the component Young's modulus and V_i represents the component volume fraction. Correspondingly, one can show that for the transverse stiffness (E_2) in Fig. 3a, b:

$$\sigma_c = \sigma_m = \sigma_f, \tag{25}$$

a condition of isostress for which one can write

$$E_2 = E_c = E_f E_m / (V_f E_m + V_m E_f), \tag{26}$$

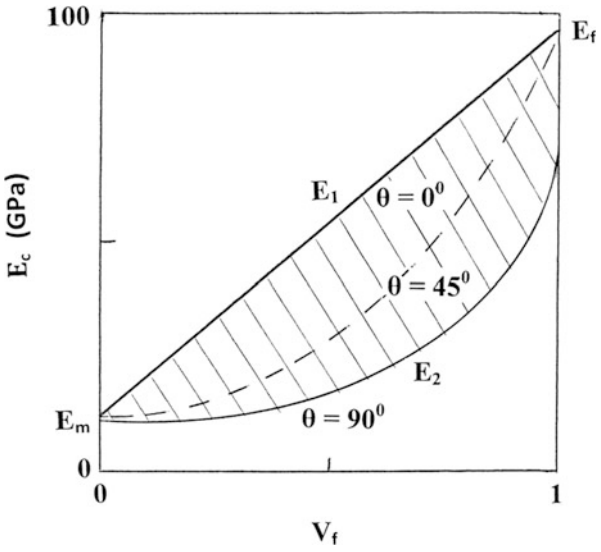
also referred to as the Reuss model (1889) of uniform stress and expressed generally by

$$\frac{1}{E_c} = \left(\frac{V_1}{E_1} + \frac{V_2}{E_2} + \frac{V_3}{E_3} + \ldots \right). \tag{27}$$

It can be observed that the rule of mixtures formula in Eq. 22 or the Voigt model in Eq. 24 for a 2-component or 2-phase system having $E_f >> E_m$ constitutes an upper boundary for composite stiffness relative to the Reuss model in Eq. 27, which for a 2-component system represented by Eq. 26 forms a lower boundary for composite stiffness which can be represented generally in Fig. 4 for volume fractions (V_f) of stiff fibers ($E_f >> E_m$). We can see from Eq. 21 that the composite stiffness at 45° (Fig. 3c, d) is closer to the lower boundary as shown by the dotted curve in Fig. 4 and trending toward the lower boundary as θ increases at 90° (Fig. 3c).

Figure 5a compares the general trend for 2-phase composite stiffness variations for continuous (longitudinal) fibers in contrast to discontinuous, directional whiskers having $l/d > 10^2$ and $l >> l_c$, as well as particles where the reinforcement is noticeably reduced at volume fractions <10 %. In addition, Fig. 5b shows that in general, continuous fiber diameters are not of any consequence in the context of fiber volume fraction, confirming the applicability of the rule of mixtures formula in Eq. 22. This can be applied generally in the so-called Voigt (1889) average:

$$P_c = \sum_{i=1}^{n} P_i V_i \tag{28}$$



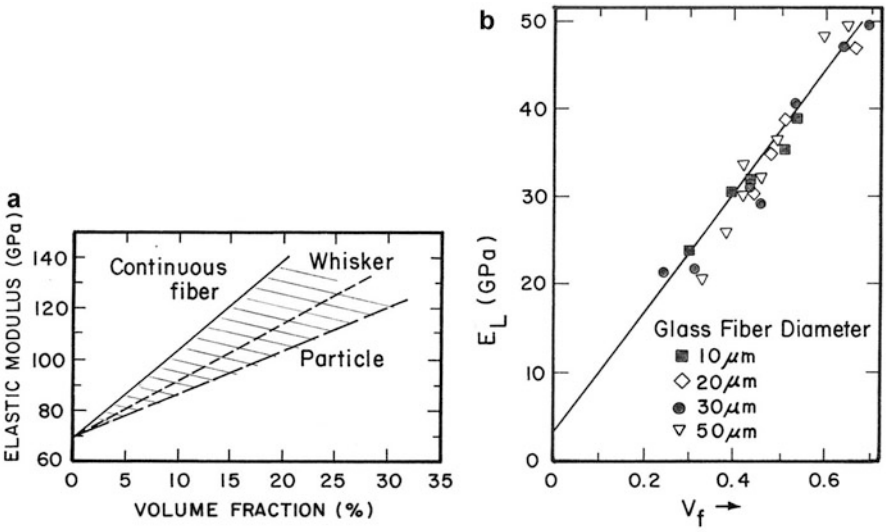


Fig. 4 Upper boundary (E_1: uniform strain) and lower boundary (E_2: uniform stress) composite stiffness versus fiber volume fraction. $E_f >> E_m$

Fig. 5 (a) Comparison of composite stiffness (modulus) versus volume fraction (V_f) for different reinforcing phase forms. (b) Composite stiffness (E_L) or longitudinal elastic modulus as a function of fiber volume fraction (V_f) for various diameter glass fibers in epoxy. Adapted from Meyers and Chawla (1984)

where P represents a property (such as strength, thermal expansion coefficient, etc.) of the composite (c) and the ith component having a volume fraction V_i.

Similarly, the Reuss (1929) *model* implies that

Table 1 Properties of some typical MMC and PMC composites in the longitudinal and transverse directions[a]

Composite	Matrix	E_m (GPa)	Fiber	Volume fraction (V_f)	Density (g/cm^3)	Elastic modulus (GPa)		Tensile strength (GPa)	
						Long.	Trans.	Long.	Trans.
MMC	Al	30	B	0.50	2.65	210	150	1.50	0.14
MMC	Ti-6Al-4V	110	SiC	0.35	3.86	300	150	1.75	0.41
MMC	Al–Li	40	Al_2O_3	0.60	3.45	262	152	0.69	0.18
PMC	Epoxy	8	E-glass (unidirectional)	0.60	2.0	40	10	0.78	0.028
PMC	Epoxy	8	2-D glass cloth	0.35	1.7	16.5	16.5	0.28	0.28
PMC	Epoxy	8	Boron (unidirectional)	0.60	2.1	215	9.3	1.40	0.063
PMC	Epoxy	8	Carbon	0.60	1.9	145	9.4	1.86	0.065
PMC	Polyester	2	Chopped glass	0.70	1.8	55–138	–	0.103–0.206	–

[a]Adapted from Schaffer et al. (1999)

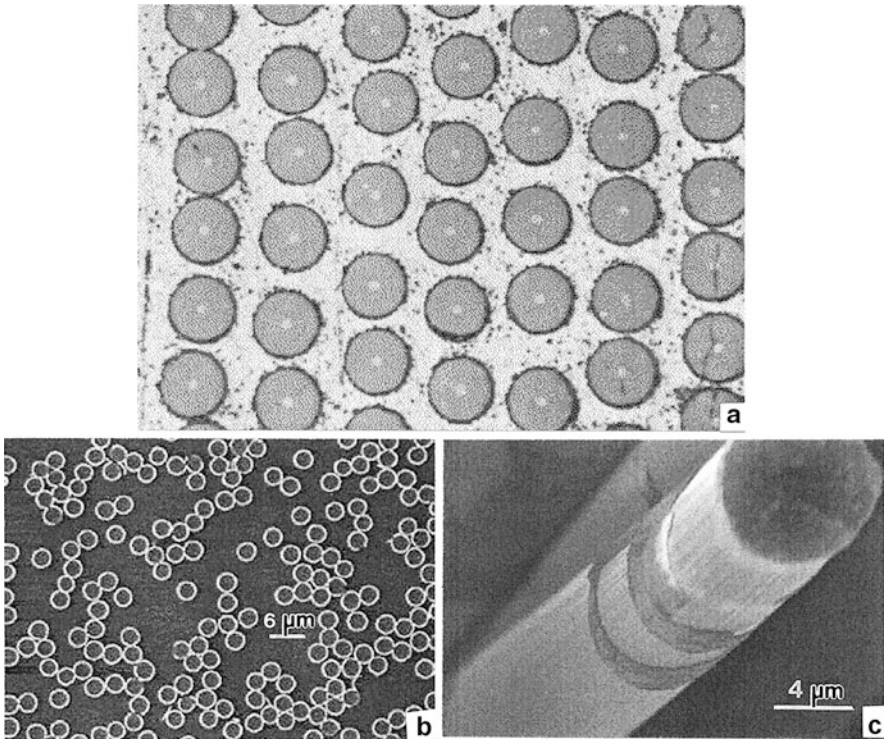


Fig. 6 (a) Al/B-fiber composite (transverse section view). The B fibers have a diameter of 100 mm and a volume fraction (V_f) of 0.44 (44%). (b) and (c) show coated carbon fibers for compatibility and reduced diffusion reactions. (b) shows an SEM image cross-section of ZrO_2-coated C fibers (oxide coating is white ring wound fibers) in a ZrC matrix. (c) shows an SEM view of ZrO_2/HfO multilayer-coated carbon fibers(Courtesy of Ultramet)

$$\frac{1}{P_c} = \sum_{i=1}^{n} \left(\frac{V_i}{P_i} \right) \tag{29}$$

for transverse composite properties, P_c and P_i.

In addition to the general classification of composites characterized by fiber or the dispersed phase in a matrix, the matrix itself becomes part of the characterization. Prominent matrices include of course metals or alloys (metal matrix composites: MMC), polymers including thermosetting epoxies (polymer matrix composites (PMCs)), and reinforced ceramics (ceramic matrix composites (CMCs)). In both MMCs and PMCs, the reinforcing phase (fibers) enhances stiffness, strength, and creep of the matrix while the principal enhancement for CMCs is toughness and the ability to sustain intrinsic strength even when cracks form.

Table 1 illustrates a few metal matrix and polymer matrix composites, while Fig. 6a shows a typical transverse plane view for an aluminum/boron fiber

Table 2 Properties of fibers[a]

Material class	Material	E (GPa)	T.S. (GPA)	P(g/cm^3)	Specific stiffness E/ρ	Specific strength T.S./ρ
Metals	Be	315	1.3	1.8	175	0.72
	Pearlitic steel	210	4.2	7.9	27	0.53
	Stainless steel	203	2.1	7.9	26	0.27
	Mo	343	2.1	10.3	33	0.20
	β-Ti	119	2.3	4.6	26	0.50
	W	350	3.9	19.3	18	0.20
Inorganics	Al_2O_3	470	2.0	3.96	119	0.51
	Al_2O_3 whiskers	470	2.0–20.0	3.96	119	–
	B	385	7.0	2.6	148	2.69
	BN	90	1.4	1.9	47	0.74
	Graphite	490	3.2	1.9	258	1.68
	Graphite (Kevlar)	133	2.8	1.5	89	1.87
	E-glass	84	4.6	2.55	33	1.80
	S-glass	72	6.0	2.5	29	2.40
	Mica	231	3.2	2.7	86	1.19
	SiC	380	2.8	2.7	141	1.04
	SiC whiskers	470	2.0–20.0	3.17	148	–
	Si_3N_4	380	1.0–10.0	3.8	100	–
Organic	Nylon 66	4.9	1.05	1.1	4	0.95

[a]Adapted from Courtney (1990)

composite corresponding roughly to the example provided in Table 1. Table 2 shows some properties for fibers, metallic, inorganic, and organic. The tensile strength for the fibers is noted by T.S., which corresponds to the ultimate tensile stress. Fibers and often other composite systems are also characterized by specific stiffness, the stiffness E divided by the density, ρ, and the specific strength, tensile strength (T.S.) divided by the density: T.S./ρ, where in some cases T.S. $= \sigma_y$, the yield stress. In many cases, the specific strength is referred to as the strength-to-weight ratio, and for many applications, strategies to increase this ratio in a significant way can indicate notable performance enhancement.

We should comment on the fact that in some cases of matrix reinforcement using continuous or long fibers, the fiber may be incompatible with the matrix forming a weak bond (or low interfacial free energy or adhesive energy) or conducive to diffusive phenomena at the fiber/matrix interface creating a reactive phase or a new interfacial chemical phase. This can cause fiber pullout, rendering the

reinforcement seriously compromised, especially over time and at elevated temperatures promoting interfacial diffusion and reaction. Figure 6b, c illustrates the coating of commercial fibers to enhance fiber/matrix compatibility and the elimination of reaction.

References

Chawla KK (1987) Composite materials science and engineering. Springer, New York
Courtney TH (1990) Mechanical behavior of materials. McGraw-Hill, New York
Kelly A (1976) Chapter 9: Composites with brittle matrices. In: Murr LE, Stein C (eds) Frontiers in materials science. Marcel Dekker, New York, pp 335–364
Meyers MA, Chawla KK (1984) Mechanical metallurgy: principles and applications. Prentice-Hall, Englewood Cliffs
Reuss A (1929) Z Berechnund der fliebgrenze von misch-kristallen fut grund der plastizitats-bedingung for ein kristalle. Angew Math Mech 9:49–55
Schaffer JP, Saxena A, Antolovich SD, Sanders TH Jr, Warner SB (1999) The science and design of engineering materials. McGraw-Hill, New York
Voigt W (1889) Weid. Ann, 38; Lehrbuch der Kristall Physik. Leipzig, Teubner (1910)

Eutectic Composites

Contents

Abstract

Examples of eutectic fibers formed in a metal matrix by directional solidification are presented.

Introduction

It has been nearly a half century since novel eutectic composites have been created by directional solidification of binary alloy systems. In this short chapter a few examples of these 2-phase composites are examined.

Narrative

Eutectic composites are essentially the essence of the nickel-base superalloys described in part 6, chapter "▶ Classification of Composite Materials and Structures." ▶ Figure 9 in chapter "Classification of Composite Materials and Structures" illustrates the creation of fcc eutectic $Ni_3(Al, Ta)$ blocks surrounded by an fcc Ni-Cr solid-solution matrix. In the development of such composite systems, the goal has involved the creation of a multifunctional material. For example, Ni-base superalloys have compositions which promote prominent corrosion (oxidation) resistance, high creep strength or creep-rupture strength, high-temperature resistance, good fracture toughness, etc. These properties or performance functions are either shared between the phase compositions or are

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_24

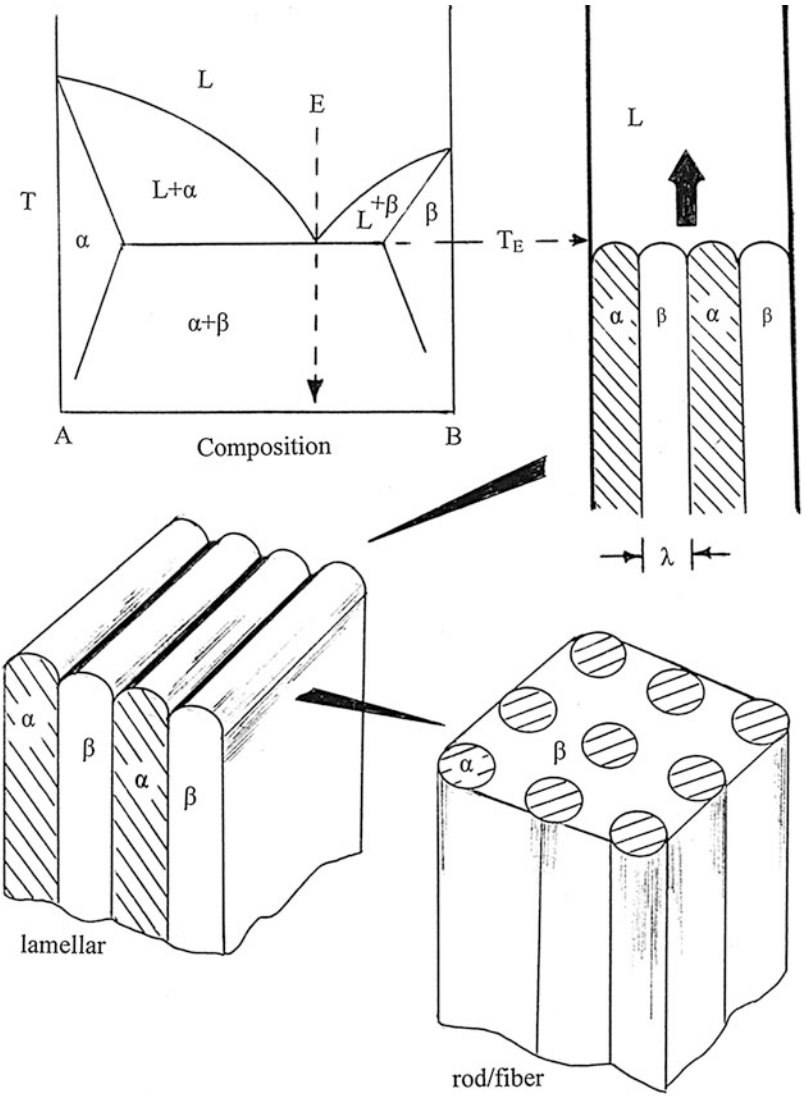


Fig. 1 Eutectic (binary) phase diagram and associated unidirectional solidification creating 2-phase composite alloy. E denotes eutectic at temperature T_E. L is liquid (From Murr et al. 2012)

distinct properties or functions exhibited by a phase. Eutectic phases are normally very hard and as a consequence are brittle or have reduced fracture toughness. Consequently, there is a compatibility advantage for having a hard, strong phase nested in a ductile environment such as the γ/γ′ system shown in Fig. 9 where the γ-matrix initially accommodates the accumulation of strain by building dislocation substructures which are eventually transferred to the γ′-phase which continues to strengthen the composite. In these systems, the second phase is more a

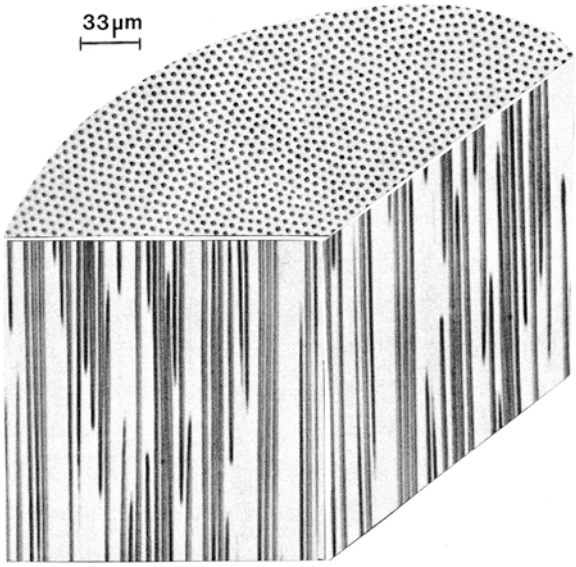


Fig. 2 3-D optical metallographic composition for aligned MnSb eutectic rods in Sb matrix where l/d > 10 (From Murr et al. 2012)

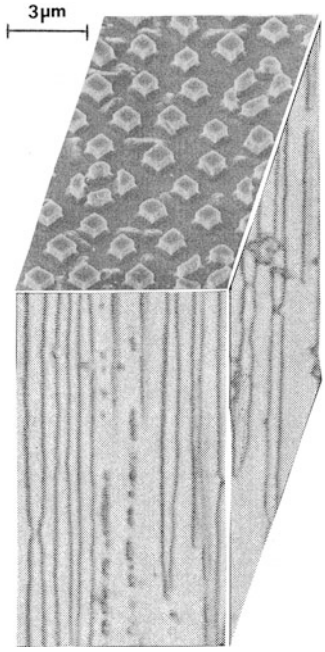


Fig. 3 3-D image composition (SEM horizontal plane and optical metallography in the vertical planes) for Ni_3Al eutectic rods in a Mo matrix. $l/d > 10^2$

strengthening agent according to Eqs. 1 and 28 rather than providing significant stiffness enhancement.

As implicit in Fig. 24a, fibers, particularly long fibers, provide significant reinforcement of nominally ductile matrices. To achieve this feature in eutectic

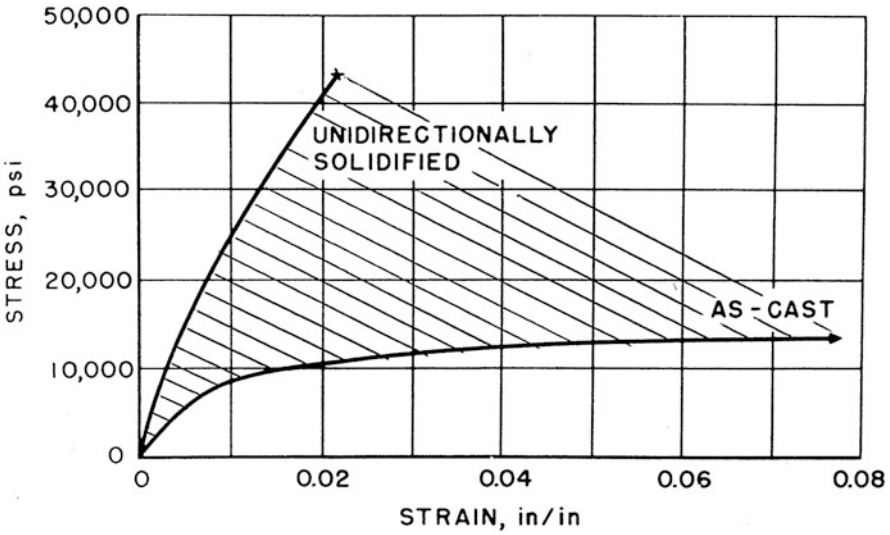


Fig. 4 Stress–strain diagrams for directionally solidified and as-cast Al-Al_3Ni eutectic structures. 10,000 psi = 10 ksi = 0.069 GPa (Adapted from Hertzberg 1976)

alloys involves directional solidification which transforms the liquid alloy into two distinctly different phases: a rod or filament-like eutectic phase dispersed in a solid matrix or a solid-solution matrix involving a binary composition different from the eutectic phase composition as illustrated in Fig. 1.

In the fabrication of eutectic-reinforced matrix composites, the unidirectional solidification of an alloy ingot of eutectic composition is achieved under a steep axial thermal gradient by gradual withdrawal of an ingot from a furnace. This establishes a planar solid–liquid interface separating the liquid (melt) and solid phases shown by the large directional arrow in the liquid (L) in Fig. 1 (upper left). The two solid phases (α and β) produced from the eutectic reaction at the eutectic temperature (T_E in Fig. 1) grow parallel to the direction of solidification (large arrow in Fig. 1). Hertzberg (1976) has illustrated that the type of morphology developed in a eutectic microstructure depends on the relative volume fraction of each phase. Eutectic rods or filaments embedded in a matrix phase will form when one phase is less than that $1/\pi$ of the total alloy volume. These features are illustrated in Fig. 1. Figure 1 also illustrates that the spacing, λ, of lamellar phase structures in particular is controlled by the velocity of the solid–liquid interface:

$$\lambda = \sqrt{\frac{C}{R}} \tag{1}$$

where C is a constant and R is the solidification rate. Consequently, controlling R for any specific phase composition (diagram) as shown in Fig. 7 will allow the

relative volume fraction and microstructural morphology to be controlled as implicit in Fig. 7.

Figures 2 and 3 illustrate the fabrication of eutectic filaments or rods in supporting (or reinforced) matrices. Figure 2 shows MnSb eutectic rods in an hcp-Sb matrix, while Fig. 3 shows a 3-D composition illustrating Ni_3Al (γ'-fcc) eutectic filaments in a Mo (α-bcc) matrix. Figure 4 illustrates the strengthening for directionally solidified Al-Ni creating Al_3Ni fibers in an Al matrix in contrast to a cast Al-Ni composition. This strengthening is consistent with that achieved for B fibers in Al indicated in Table 1, where the Al tensile strength is roughly 0.4 GPa in contrast to a B-reinforced Al matrix where the strength is ~1.4 GPa.

References

Hertzberg RW (1976) Structure-property relationships in eutectic composites. In: Murr LE, Stein C (eds) Frontiers in materials science. Marcel Dekker, New York, pp 421–454

Murr LE, Martinez E, Gaytan SM, Ramirez DA (2012) Contributions of light microscopy to contemporary materials characterization: the new directional solidification. Metallogr Microstruct Anal 1:45–58

Examples of Natural Composites and Composite Structures

Contents

Abstract

Natural or biological composites are reviewed with a number of appropriate examples: human bone and teeth, wood, pearls, mother of pearl, and related shell structures.

Introduction

There are many examples and variances of composite systems in the natural or biological world. If one stands before a mirror, the ridges characterizing the hydroxyapatite crystal reinforcements can be visualized, and their effect on the strength of the teeth further appreciated on biting a piece of a hard candy. The same appreciation accrues on examining a piece of wood where its properties are readily associated with the orientation of its grain.

After exploring some of the composite fundamentals associated with biological materials or materials systems in this chapter, more specific applications will be explored in chapter "▶ Structure and Function of Viruses and Bacteria" which follows.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_25

Human Bones and Teeth

Mc Kee et al. (2005) have described bones and teeth as heterogeneous, hierarchical, composite structures with distinct functions. Bone components include a crystalline (nanocrystalline) hydroxyapatite phase, a similar composition amorphous phase, a crystalline organic phase, collagen, and protein molecules forming an amorphous, organic phase in the form of gels and sols, and other liquids. The largest or longest human bone, the leg femur, consists of an outer hard, porous, or foam-like cortical (or cancellous) bone having Young's modulus ranging from ~18 to 20 GPa surrounding a narrow cavity consisting of struts of trabecular (cancellous) or softer bone having a Young's modulus ranging from ~0.5 to 2 GPa. Figure 1 illustrates these structural or macrostructural features.

At the tissue level, bones consist of cells composing bone tissue and vascular tissue cells, various organic matrices, and an inorganic matrix composed primarily of hydroxyapatite: $Ca_{10}(PO_4)_6(OH)_2$ having a hexagonal crystal structure, a = b = 9.5 Å, c = 6.9 Å, as well as a chemically similar amorphous phase. The organic matrix of the bone produced by cells in the extracellular matrix forms collagen molecules or macromolecules (tropocollagen) consisting of three amino acid-related polypeptide molecular chains wound into repeating, triple-helical collagen macromolecules ~300 nm long and 1.5 nm in diameter. These can form repeating fibril arrays that can accommodate a variety of shape requirements and provide tissue flexibility. The repeating arrays of collagen macromolecules and fibrils illustrated in Fig. 2a, b are arranged in a staggered configuration creating fibrils having an observable periodicity known as D-band structure where D = 67 nm. This structure includes a gap-like region. The fibrils shown in Fig. 2, consisting of overlapping tropocollagen molecules which are cross-linked covalently to each other using lysine side chains, only occur in collagen. The fibril diameters represented in Fig. 2b can vary from 30 nm in the cornea to ~500 nm in tendons, where collagen fibrils (Fig. 2b) bundle to form larger fibers aided by cross-linking macromolecules such as proteoglycans. Collagen molecules or fibrils (Fig. 2b) are the most abundant protein construction material in the human body. There are actually 27 different collagen types which can form complex, branched structures in the extracellular, noncollagenous matrix proteins.

In skin and similar tissue, collagen fibrils are aligned parallel forming sheet-like structures as illustrated in Fig. 3. Ligaments and tendons, as noted, have collagen fibers or fiber-matrix structures. However, in bone, the formation of hydroxyapatite nanocrystals is controlled by osteoblast cells which, when engulfed in hydroxyapatite crystals, become osteocytes. A third cell type called osteoclast cells removes bone mineral and bone matrix and regulates the formation and resorption (or turnover) of bone. Hydroxyapatite crystals form as nanocrystals 30–150 nm long, 15–30 nm wide, and 2–10 nm thick, with the long dimension oriented in the c-axis [001] direction (McKee et al. 2005; Glimcher 2006). These crystals, or rather their formation, are controlled by the collagen matrix and associated gene structure of proteins. These gene structures signal the nucleation and growth as

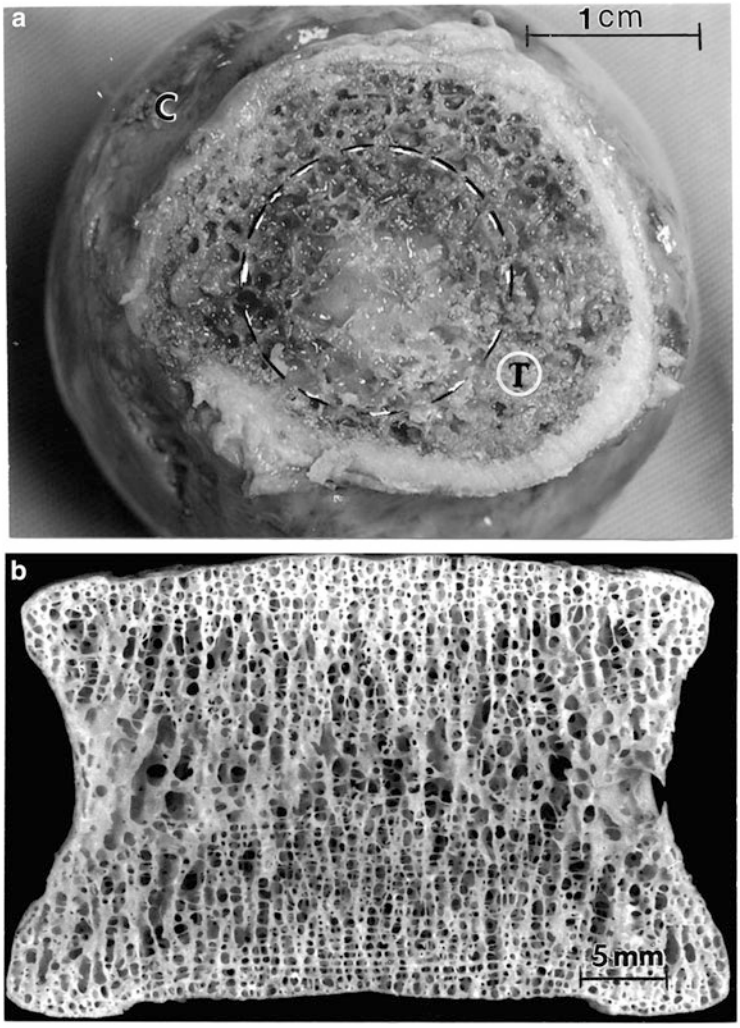


Fig. 1 Human bone structure. (**a**) *Upper* femur showing outer cortical (C) region and softer, inner trabecular (T) bone region or intramedullary regime (*dotted*). (**b**) Cortical bone section or longitudinal view of (**a**) showing porous bone structure (**b**) is courtesy of Prof. Alan Boyde

well as site-specific attachment, size, and density of the hydroxyapatite crystals which orient parallel to the collagen macromolecules as illustrated in Fig. 2c.

The crystal deposition on the collagen fibril scaffold produces configurations which allow the collagen fibrils to bear compressive loads, while maintaining flexibility. Changing the aligned hydroxyapatite crystal density and size through functional signaling by the cell structure can change the strength of the composite. This is especially notable in the structure of teeth (or tooth dentin and enamel) where the hydroxyapatite crystals are $\sim$10 times larger than bone in all dimensions,

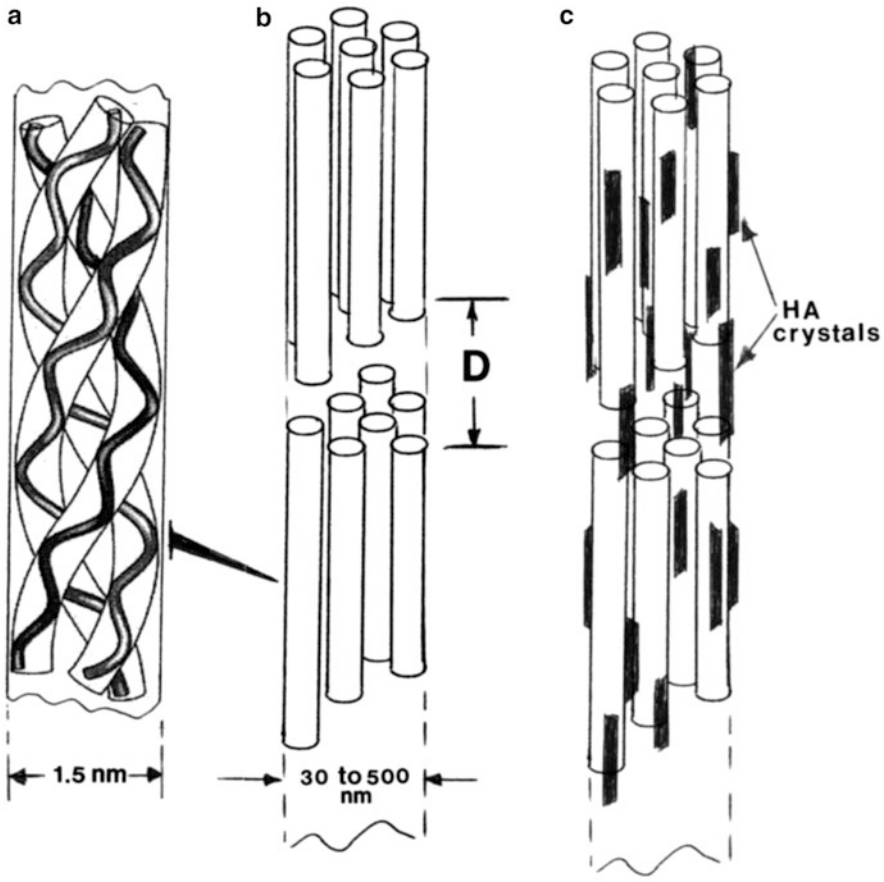


Fig. 2 Schematic views of hierarchical molecule and fibril structure. (**a**) Helical triple polypeptide chains forming a tropocollagen molecule (**a**). These long-stranded tropocollagen molecules are covalently cross-linked using lysine side chains to form bundles or collagen fibrils (**b**). The staggered fibrils form gaps, called D-bands or D-periods. Hydroxyapatite (HA) crystals can form in these gaps and on tropocollagen scaffold surfaces (**c**)

providing a more dense and much stronger, oriented composite structure (Kirkham et al. 1998) as illustrated in Fig. 4. This is especially true for enamel which is roughly 95 % hydroxyapatite. Ideally, this implies that the volume fraction of oriented hydroxyapatite crystals strengthens the interwoven collagen matrix according to Eq. 22 and as represented more generally in ▶ Fig. 4 (chapter "Eutectic Composites"), where Young's modulus for collagen can be approximated from tensile testing of tendons to be ~150 kg/mm^2 (1.47 GPa) (Piekarski 1973) and hydroxyapatite.

Young's modulus has been measured to vary from ~1,755 to 17,600 kg/mm^2 (17.2–172 GPa) (Mack 1964; Currey 1964), a factor of roughly 10–100 times

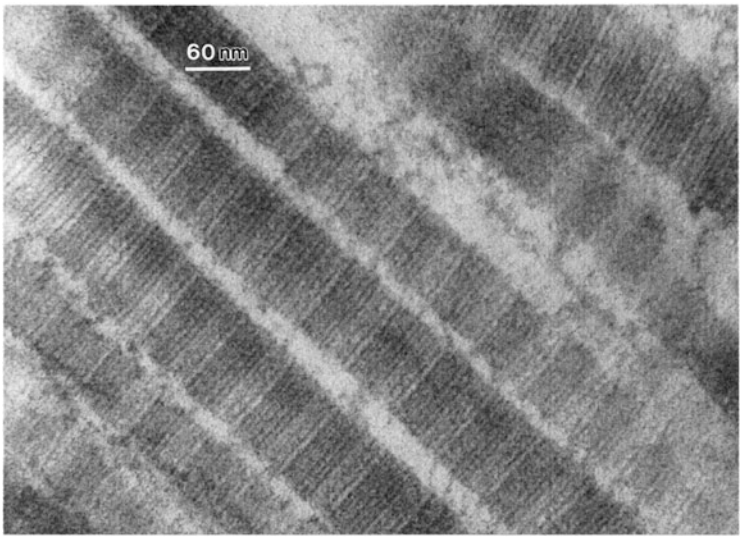


Fig. 3 TEM image showing collagen fibrils with D-band structure for a thin section through mammalian lung tissue (Courtesy of Louisa Howard). These fibrils have no associated hydroxyapatite crystals because they compose soft tissue. As the tissue function changes in the context of strength requirements, the biological (protein) cell signaling will direct hydroxyapatite nucleation and growth accordingly. The collagen lengths are also shorter than in tendon or bone structures

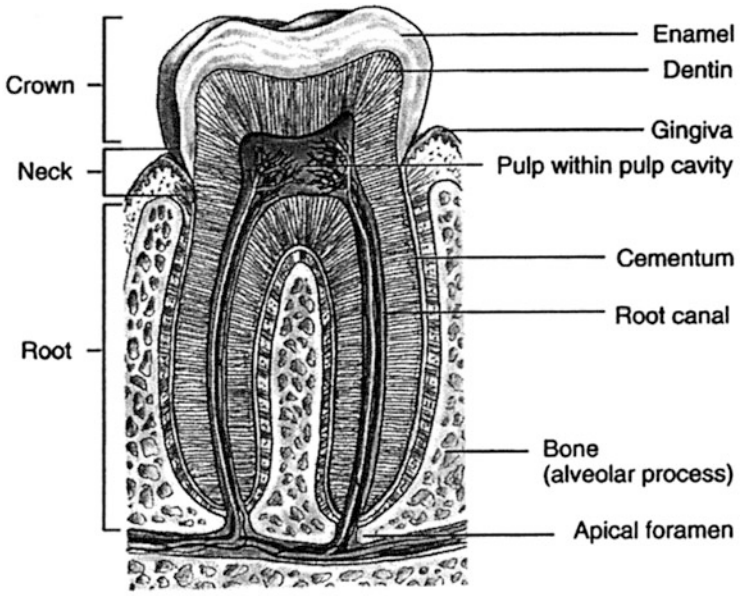


Fig. 4 Composite structure of a human tooth showing functional components

greater than the organic matrix: E_f (hydroxyapatite) ≅ 10–100 E_m (collagen matrix). However, in the case of porous solid or cellular foam structures, Young's modulus or stiffness (E) is given by Gibson and Ashby (1982):

$$E = E_o\left(\frac{\rho}{\rho_o}\right)^2 \tag{1}$$

where ρ/ρ_o is the relative density (ρ is the porous, foam density while ρ_o is the full-solid density) and E_o is fully dense Young's modulus. Consequently, the stiffness of hard (cortical) bone (Fig. 1b) (∼18–20 GPa) reflects this porosity or density variation of bone. Cortical bone porosity is ∼62 %. It can be noted that porosity and density (or relative density) are related by

$$\text{porosity}(\%) = \left(1 - \left(\frac{\rho}{\rho_o}\right)\right) \times 100 \tag{2}$$

Bone consists generally of 35–45 % hydroxyapatite crystals, 35–45 % organic matter (including of course collagen fibers), and 15–25 % water. This is in contrast to roughly 75 % collagen in skin and other tissues (Fig. 3).

The structure of the teeth (Fig. 4) is dominated by dentin and enamel composing the tooth crown. The enamel is the hardest of all body tissue and is composed by hydroxyapatite rods (4–8 μm diameter) or prisms consisting of crystal bundles and roughly 2 % collagen matrix and ∼4 % water. These rods run from the dentin/enamel interface and while generally perpendicular to the dentin/enamel interface, they are also intertwined and interlocking in a complex arrangement as shown in Fig. 5a. The composing hydroxyapatite crystals are roughly 160 nm long and 20–40 nm wide. The enamel stiffness or Young's modulus is roughly 83 GPa (Staines et al. 1981).

The underlying dentin regime in a tooth (Fig. 4) forms the main mass of the tooth and supports the enamel. It absorbs the pressure applied to the tooth in a composite similar to bone, although containing tubules parallel to the collagen fibrils and fibers containing hydroxyapatite crystals as in Fig. 2c, in varying amounts ranging from a higher concentration at the enamel/dentin interface and lesser amounts toward the root areas and areas interfacing with bone. Figure 5b illustrates a section of dentin near the enamel/dentin interface. As the dentin regime moves away from the dentin/enamel interface, the tubules increase in density. That is, the porous volume increases. These structures are similar to a honeycomb structure to be described later, especially in wood structure. Corresponding dentin Young's modulus is roughly 12 GPa (Povolo and Hermida 2000) which, in contrast to the enamel as indicated above, is a factor of 7 less, although it is greater near the dentin/enamel interface, but changes with hydroxyapatite crystal content as in bone (Ferguson et al. 2005). The collagen structures and their role in both bone and tooth function are also an important feature (Fratz 2008).

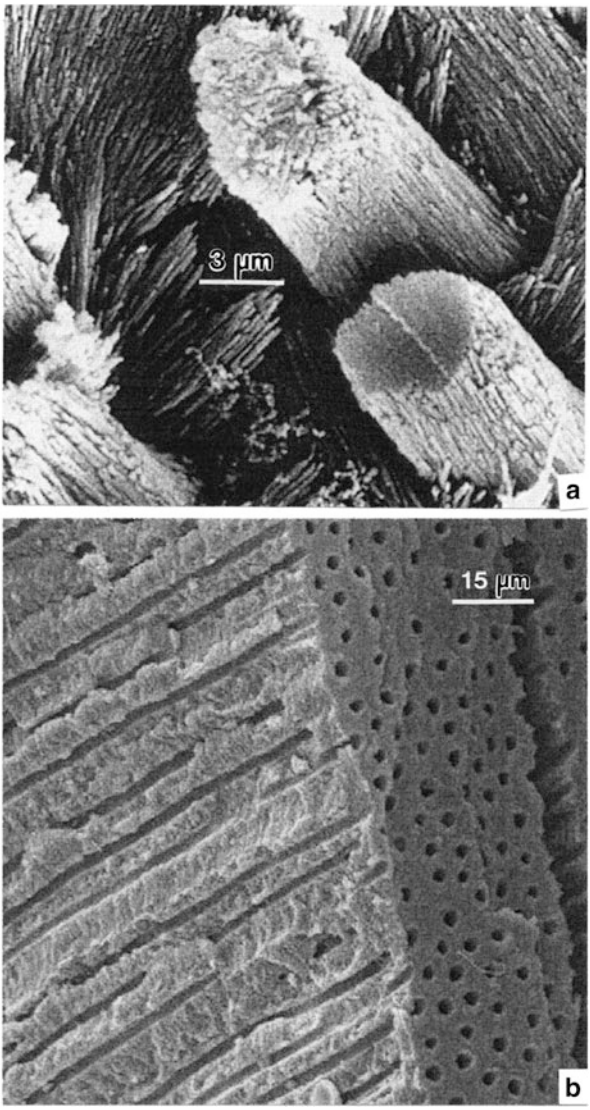


Fig. 5 Tooth composite regimes. (**a**) Hydroxyapatite crystal rods and crystal fibrils in tooth enamel. (**b**) Dentin structure showing tubules which contain water and cellular material

Mollusk Shells and Cultured, Freshwater Pearls

Like bone and teeth, shell-bearing mollusks, including oysters and mussels, and their included pearls, exhibit novel, organic/inorganic composite structures similar in some respects to bone and teeth, although the inorganic, crystalline phase is calcium carbonate (in one or the other allotropic crystal structures: calcite or aragonite) rather than hexagonal hydroxyapatite. There are more than 10^5 different shell-bearing ocean or freshwater mollusks (or mussels). Of these, there are roughly

10^3 bivalve mussel species of which one-third occupy freshwaters, while two-thirds occupy salt waters. These two-valve species produce pearls in the valve body, or mantle, by enveloping a foreign particle in an organic, molecular sac which then mediates the formation of a complex calcium carbonate/organic forming a strong and hard encasement.

Similar to bone and teeth, mussel shells and pearls are formed by a macromolecular (cellular) framework which controls the formation of calcium carbonate crystals: their nucleation, site specificity or positioning on the organic layer, size, and polymorph, either $CaCO_3$ calcite (trigonal crystal structure: a = 5Å, c = 17.1 Å) or aragonite (orthorhombic crystal structure: a = 5Å, b = 8Å, c = 5.7Å). These crystals can form as polygonal tiles or plates separated by the interlamellar layers of organic matrix, formed by macromolecules such as polysaccharide β-chitin, a hydrophobic silk protein called conchiolin, and a complex assemblage of other hydrophilic proteins (Lowernstam and Weiner 1989). These layers signal the formation of the crystal platelet which can be oriented with the longest lattice parameter parallel or perpendicular to the organic interlamellar layer, which in many cases is discontinuous. This is a rather remarkable process in the context of crystalline materials nucleation and growth since in most cases crystals grow by atoms organizing epitaxially on a crystal (oriented) substrate where continued replication of this oriented growth (by layers) is energetically favorable. The process is similar to gene specification which produces and deposits hydroxyapatite crystals in bone and teeth.

As illustrated in Fig. 6, the structure of the abalone shell consists of outer, dense calcite crystal multilayers oriented in the c-axis direction, similar to the dense enamel layers in teeth (Fig. 4), with an inner layer structure consisting of aragonite tiles roughly 400 nm thick and organic interlamellar layers roughly 40 nm thick forming the nacre or mother-of-pearl structure in the shell interior (Fig. 6b). This complex, transverse, crystal/organic composite structure is extremely hard and with compressive strengths of ~0.5 GPa or higher. The composite architecture illustrated in Fig. 6c recently inspired programmable, multidirectional property development in materials systems using magnetic fields to build similar architectures in thin coatings of aluminum oxide (Erb et al. 2012).

Freshwater pearl production worldwide relies mostly on a triangular, bivalve mussel (*Hyriopsis cumingi*) slightly smaller than the abalone but with a structure similar to that shown in Fig. 6c. The valve mantle tissue in this mussel is seeded with a 3 mm square cut from shell mantle tissue illustrated in cross section in Fig. 5c. The organism responds to this irritant implant by forming a cystic tissue sac which surrounds it. This tissue receives the coded signal information for replication of the composite structures from the seed and signals the formation of rings of mostly calcite crystals and organic around the sac. These concentric composite structures propagate outward when, at some point, the macromolecular, interlamellar layer signals the formation of aragonite tiles which continue to form concentrically, accommodating the pearl radius and forming the iridescent pearl surface structures. This is illustrated in a freshwater pearl cross section near the pearl surface as shown in Fig. 7.

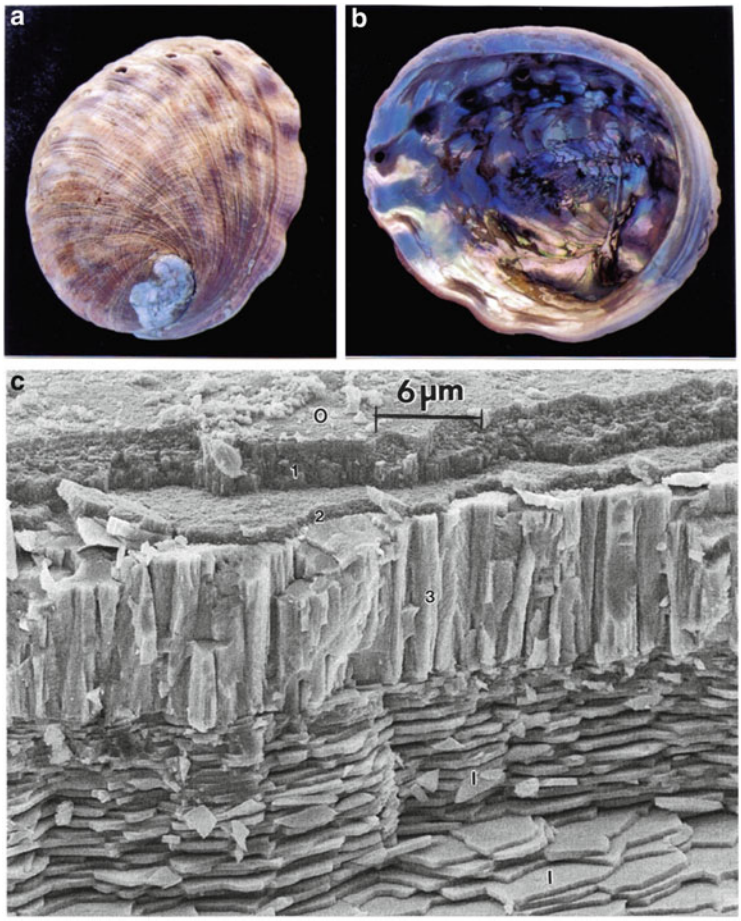


Fig. 6 Abalone (*Haliotis* sp.) shell structure. (**a**) and (**b**) Abalone shell. (**a**) is outer (calcitic) shell (O); (**b**) is inner (aragonitic) shell (I) showing iridescent mother of pearl. The long dimension of the shell is ~22 cm. (**c**) SEM view of shell cross section showing three calcite crystal layers growing in the direction of the *arrows* and multilamellar aragonite/organic layers perpendicular to the calcite orientation. The outer (O) and inner (I) (or mother-of-pearl) regions are noted ((**b**) is courtesy of ETH-Zürich)

It can be noted in Fig. 7 that this tiling of aragonite plates separated by a thin (∼40–50 nm) macromolecular (conchiolin) layer is essentially the same as the inner, mother-of-pearl surface of a mollusk shell shown generally in Fig. 6b.

Wood and Other Plant Materials

The basic unit of wood and other plant structures is the plant cell, the smallest living matter unit capable of functioning independently. Cells manufacture proteins,

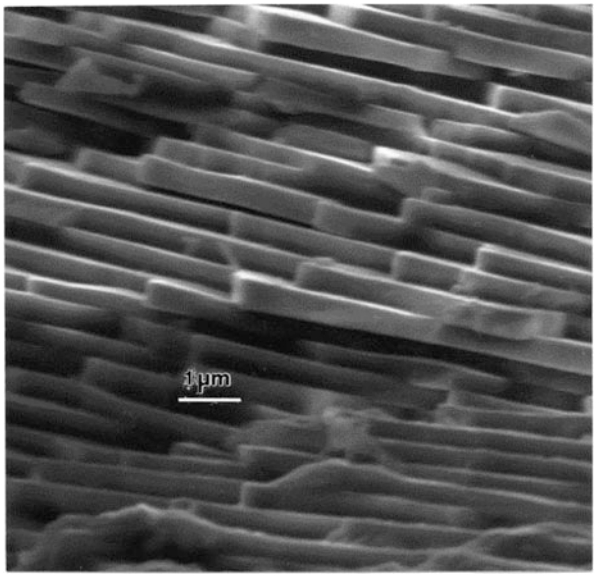


Fig. 7 SEM image showing freshwater pearl near-surface cross section of aragonite tiles ~350 nm thick (From Murr and Ramirez (2012))

polysaccharides, and mineral deposits. Polysaccharides are macromolecular carbohydrates and the principal components of wood. Polysaccharides can function as storage material, structural components, or protective regimes. Starch is the most prominent storage polysaccharide along with glycogen and other structures. Structural polysaccharides are either fibrous or matrix polysaccharides, such as the chitin-related polysaccharide forming the interlamellar structures in the aragonite tiling composing pearl as illustrated in Fig. 7.

A plant cell can vary in diameter from ~10 to 100 μm, with a cell wall outside the plasma membrane which is ~0.1–100 μm thick. The cell wall makes the cells rigid and supports the plasma membrane formed from amphipathic molecules having one end hydrophilic and one end hydrophobic. The cell nucleus contains the encryption for structure and function embedded in genetic DNA. Chloroplasts in the nucleus manufacture glucose and other sugars by photosynthesis from water and carbon dioxide. Glucose is first changed to glucose anhydride by removing one water molecule from the glucose unit which polymerizes into long-chain cellulose molecules containing from 5×10^3 to 15×10^3 glucose units. As we shall see, the basic repeating unit of the cellulose polymer consists of two 180°-rotated glucose anhydride units, often referred to as a *cellobiose unit* or cellulose monomer. Cellulose is the main structural element of cell walls, especially in wood.

Macroscopically, wood or tree stem consists of a central pith area surrounded by xylem and bark. Xylem consists of inner heartwood and outer sapwood which is surrounded by the bark sheath consisting of inner (which conducts sugars) and outer bark which serves as a protective layer. Inner bark adds new wood by the annual activity of the cambium, a layer of dividing cells sandwiched between the inner bark and sapwood. The outer sapwood functions as a water transport system which

constitutes the newly formed tree-ring region. When the sapwood region moves inward and ceases to function in water transport, it becomes heartwood.

Trees for timber production are classified as softwoods and hardwoods which are the most diverse since they contain both the lightest and heaviest timber in nature. Hardwoods are dicotyledonous angiosperms which characterize a very large group of plants including fruits and vegetables as well as weeds and trees. Softwoods belong to the more primitive plants called gymnosperms composed mostly of trees, including conifers.

Hardwoods are more structurally diverse than softwoods. In softwoods, the cells known as *tracheids* are multifunctional, providing water transport and mechanical support for the stem. In hardwoods, special, large water-transporting cells called vessels are much larger than the tracheids, and these features, along with other cellular regimes, make hardwoods more porous. Figure 8 illustrates these various cell-type structural features for hardwood sections. The vessel elements in Fig. 8a are single or double cells which are distinct from the smaller diameter, longitudinal tracheid cells called "fibers," the main cell type composing paper. In hardwoods represented by Fig. 8a, these longitudinal tracheids are ~20–30 μm in diameter and 1–2 mm in length. In Fig. 8b, the materials inside the large vessel elements are called tyloses which are ingrowths of adjoining parenchyma cells, having the structure of closed-cell, polyhedral foams, and characteristic of soft plant structure. The parenchyma cells store sugars.

In contrast to hardwood cell structure, softwood longitudinal tracheid cell structure accounts for over 90 % of the wood volume. These tracheids, illustrated in Fig. 9, are approximately 3–5 mm in length and 20–50 μm in diameter. As illustrated in Fig. 9b, this porous, anisotropic structure is similar to that of dentin illustrated in Fig. 5b where oriented collagen and hydroxyapatite crystals create a fibrous composite represented ideally in ▶ Fig. 24a (chapter "Eutectic Composites"). As illustrated in Fig. 9a, the primary cell wall layer in wood is composed of a complex arrangement of cellulose microfibrils in a matrix of hemicelluloses, lignin, and glycoproteins which are a minor constituent and thought to be involved in cross-linking. The lignin, a random, branched polymer, regulates water transport but more importantly, it stiffens the cells and the intercellular interface composed by the primary cell wall. In other plants, pectin variously substitutes for lignin. Pectin is a polysaccharide in contrast to lignin which is a complex phenolic polymer. Actually, little is known about the occurrence of pectin in wood, while it is dominant as a lignin replacement in nonwoody plants, joining cells together. When fungus enzymes break down pectin in fruits, for example, the fruit become soft and mushy.

In the cell wall structure for wood as shown in Fig. 9a, lignin is most prominent in the primary wall where it serves as an embedding and strengthening (stiffening) material for cellulose and hemicellulose. These three polysaccharide polymers, cellulose, hemicellulose, and lignin – the major biomass components in corresponding proportions of 50 % cellulose, 25 % hemicellulose, and 25 % lignin – provide the composite, structural integrity of wood or trees, which can weigh more than 2,000 metric tons. The secondary cell wall, composed of three layers,

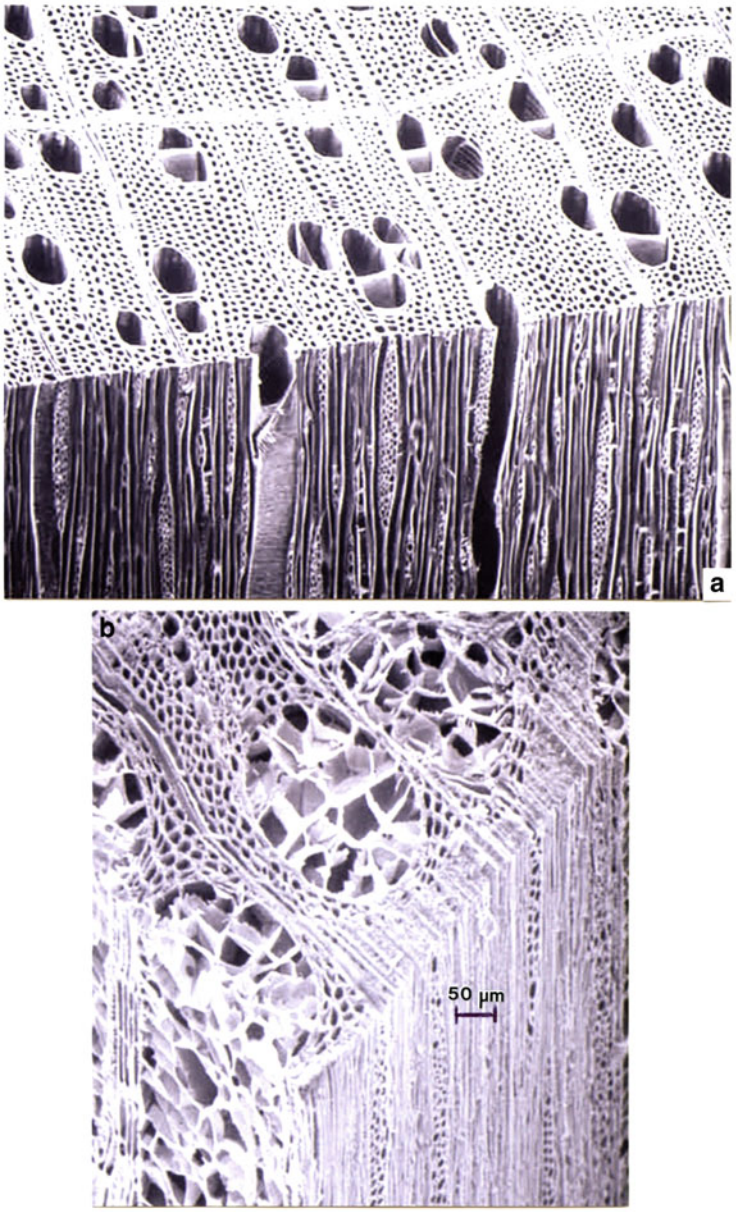


Fig. 8 Examples of hardwood structure. (**a**) Sweet birch wood block section showing vessel elements (*large holes*) and cell fibers. (**b**) Osage-orange wood block section showing tyloses in vessel elements. This is referred to as a ring-porous wood (Courtesy of the Society of Wood Science and Technology)

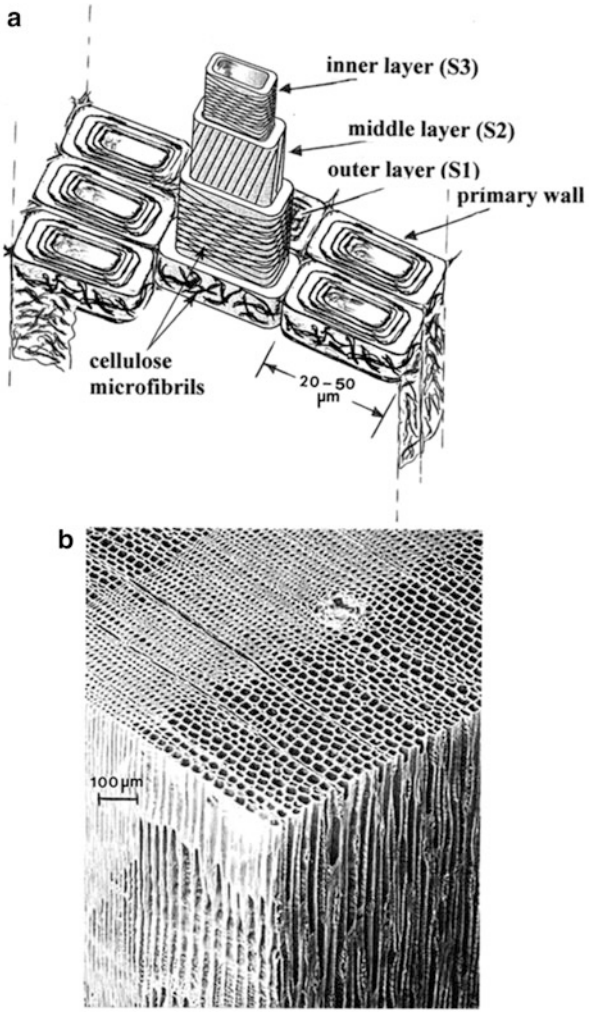


Fig. 9 Cellular grains in wood. (**a**) Cell wall structure composed of a primary layer of multidirectional cellulose microfibrils and three secondary layers with oriented cellulose microfibrils. In sapwood, the cell center contains nuclear and other matter, including chloroplasts. Higher concentrations of lignin bind the cell walls together. (**b**) Softwood (pine) structure (Courtesy of the Society of Wood Science and Technology)

S_1, S_2, and S_3, has cellulose microfibrils oriented in specific patterns. In S_1 and S_3, the microfibrils run more or less horizontally (or perpendicular to the longitudinal cell axis), while in S_2, the cellulose microfibrils run nearly parallel to the cell axis. This is in some respects characteristic of fabric weave or multidirectional laminate composite structures illustrated in ▶ Fig. 24 in chapter "Eutectic Composites."

Figure 10 illustrates the continuation of the structural hierarchy for the plant cell and especially wood cell structure shown in Fig. 9a, which is further refined in the molecular structures shown in Figs. 11, 12, and 13. In Fig. 10, the cellulose microfibril shown in Fig. 9a is shown to be composed of elementary fibrils (3–4 nm diameter) which are, as illustrated in Fig. 11a, composed of continuous cellobiose units (double glucose molecules) forming cellulose chains which are

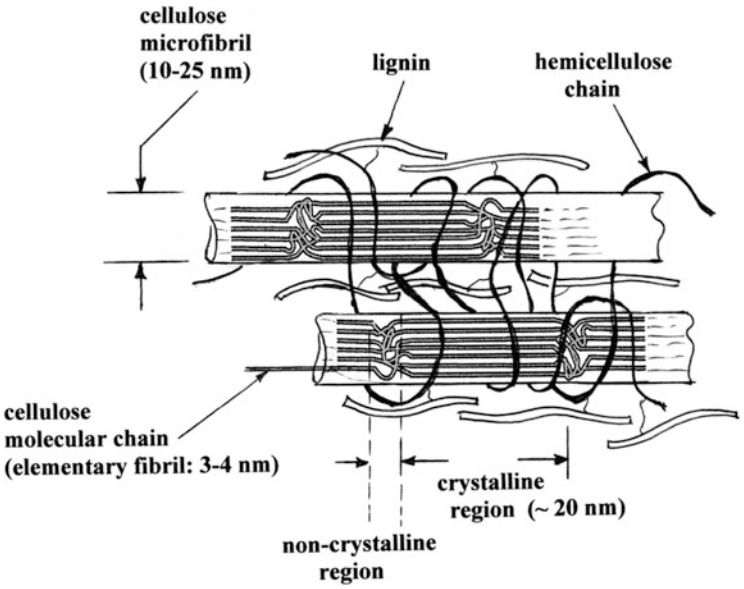


Fig. 10 Cellulose microfibril structure schematic showing hemicellulose and lignin binding

hydroxyl group (OH) cross-linked as shown in Fig. 11b. The cellobiose molecule unit is the b-axis of the monoclinic crystal unit cell with a = 0.83 nm, b = 1.03 nm, c = 0.79 nm, and $\beta = 84^\circ$ (Chakraburty et al. 2006). These cross-linked cellulose chains form crystalline or nanocrystalline regions or segments roughly 20–30 nm in length, separated by a noncrystalline, irregular chain region from bundles of elementary fibrils forming the 10–25 nm diameter cellulose microfibrils. The cellulose microfibrils are bound together by hemicellulose molecular chains which are much shorter than the cellulose chains. These molecules are characterized by glucomannans in softwoods and xylans in hardwoods (Fig. 11b). They also contain some side chains and are also bound (by H-bonds) to lignin molecular chains (Fig. 10) which are noncrystalline, high-molecular-weight polymers (protolignins) grouped into several types in plants characteristic of grasses, hardwoods, and softwoods. Figure 12a illustrates the G-lignin and S-lignin molecular units characterizing softwood and hardwood cell wall structures, respectively, although hardwoods can contain mixtures of S- and G-lignins. Figure 13 shows the more complete structure for lignin in softwood tracheids (Adler 1977). While, as noted, lignin imparts rigidity or stiffness within the cell walls (Fig. 9a), it also acts as a binding agent to hold the cells together. Like cellulose and hemicellulose, the lignin content varies in the cell wall structures in response to their functions (Brunow et al. 1998).

Table 1 illustrates these component content variances for wood and the cell wall layers, while Table 2 compares the ranges of mechanical properties for wood and cell wall components: cellulose, hemicellulose, and lignin. In the cell wall structure illustrated in Fig. 9a, the S2 layer actually composes most of the wall thickness, and

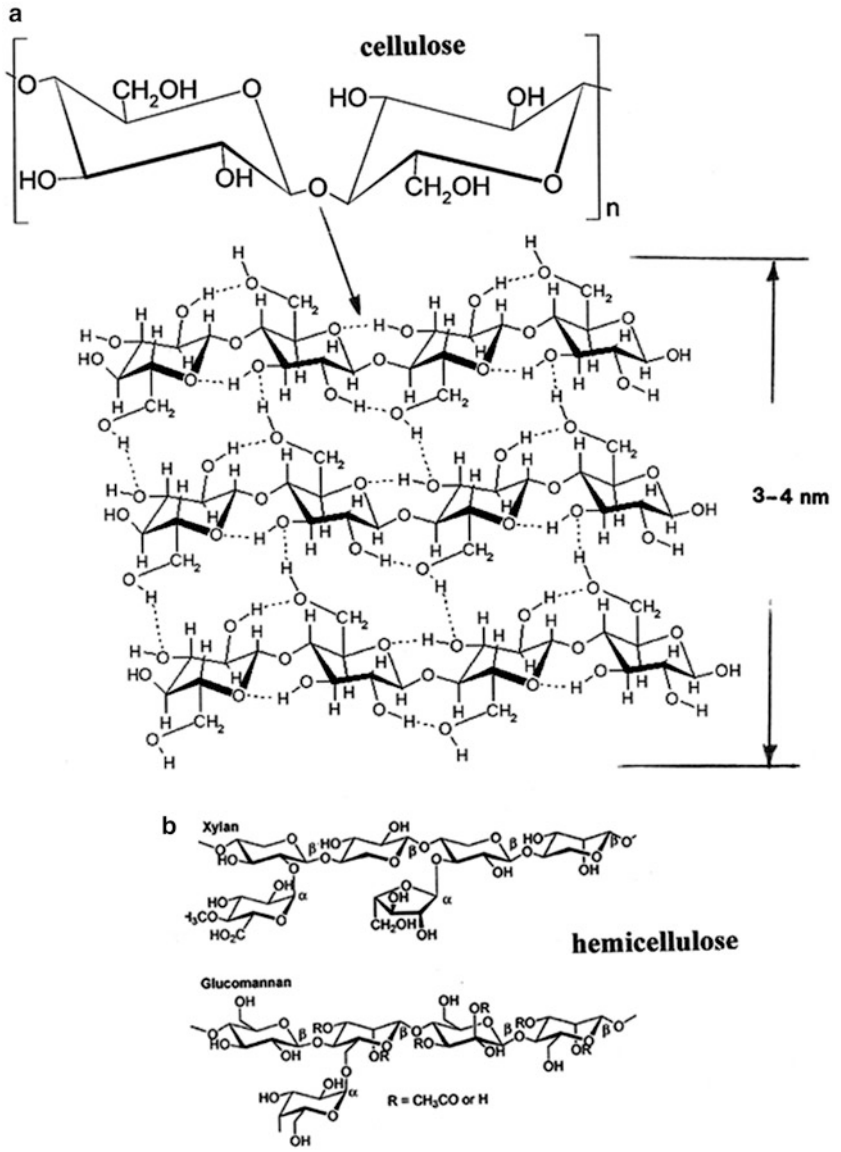


Fig. 11 Molecular structure and components for cellulose (**a**) and hemicellulose (**b**). The glucose monomer (n) is shown at the *top* and indicated by the *arrow* in (**a**)

as shown in Table 1, the composition varies through the four cell wall layers, with lignin declining dramatically from the primary layer to the outer layer (S3). As a result of the large S2 fraction and alignment of cellulose microfibrils in the S2 layer, it is a controlling feature of longitudinal modulus since Young's modulus is ~45

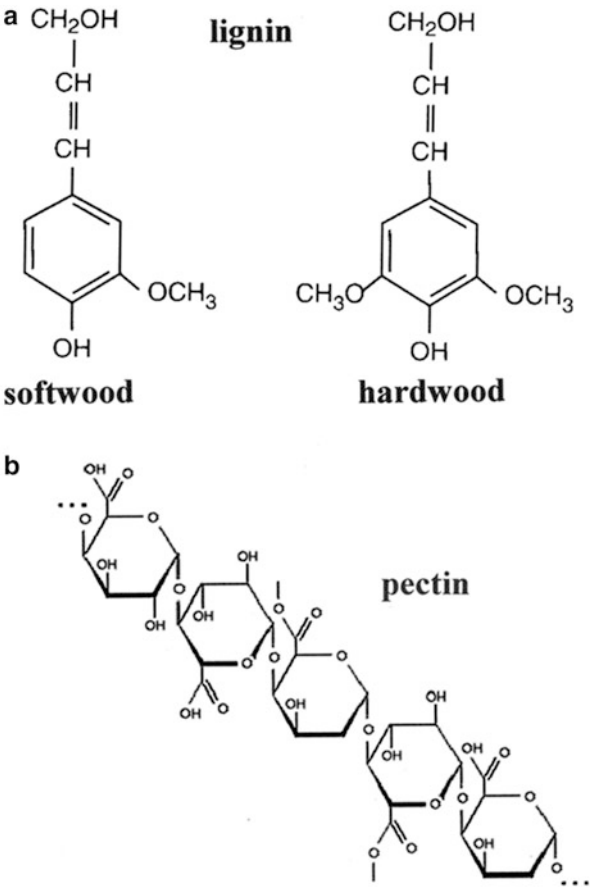


Fig. 12 Lignin units (**a**) and pectin molecule (**b**)

GPa for a microfibrillar angle of 10° in contrast to ~10 GPa at 40° (Care 1969). In contrast, the S1 layer contributes significantly to the transverse modulus as a result of the circumferential orientation of the cellulose microfibrils (Fig. 9a). More generally, the cellular structure and the cell properties of wood can be most successfully modeled as a prismatic (hexagonal) cellular honeycomb as illustrated in Fig. 15a. For loading along the grains or the longitudinal fiber direction, the cell walls compress axially and fail by buckling. Consequently, longitudinal Young's modulus (E_L) and compressive strength σ_L (along the grain) are given by Gibson et al. (2010):

$$E_L = E_{S(L)}\left(\frac{\rho}{\rho_S}\right) \tag{3}$$

$$\sigma_L = \frac{\sigma_{S(L)}}{3}\left(\frac{\rho}{\rho_S}\right), \tag{4}$$

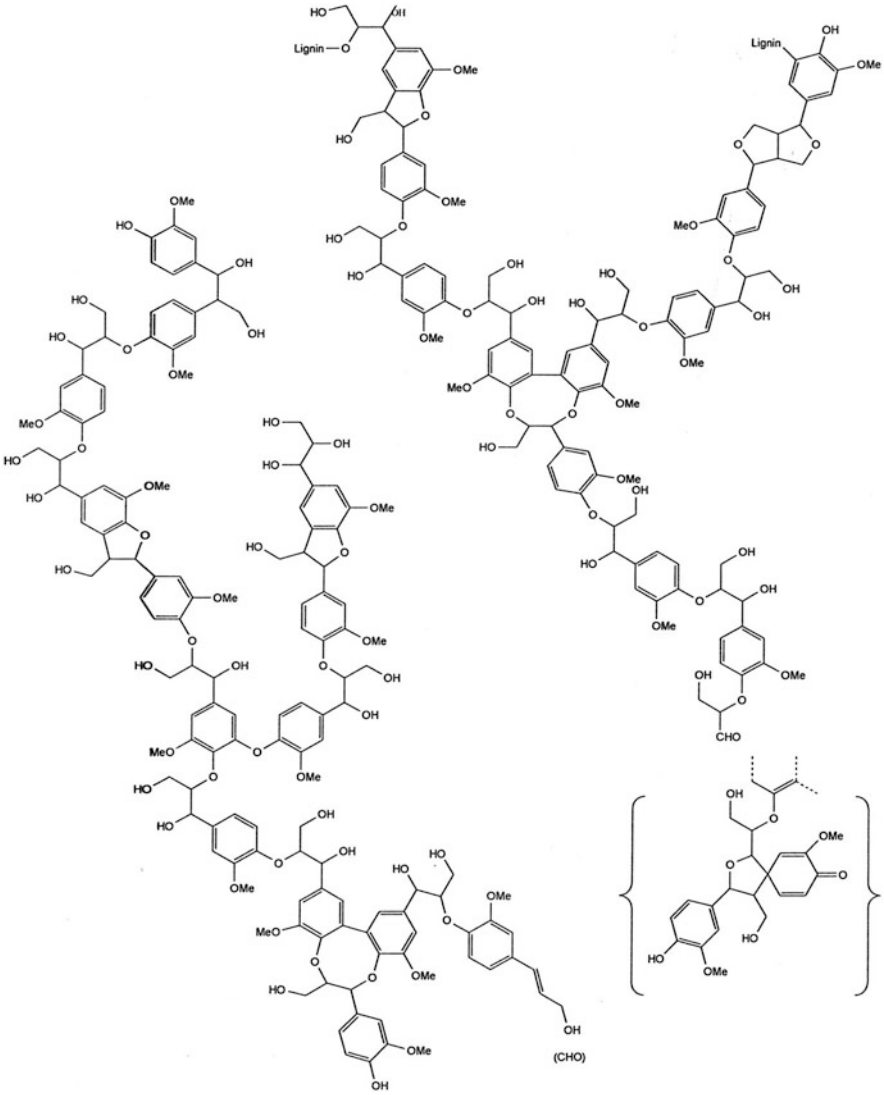


Fig. 13 Lignin (extended) softwood structure. *Me* represents the methyl group, CH_3

where the factor of 1/3 in Eq. 4 relates the difference between tensile and compressive strengths: (σ_L) tensile $\cong 3(\sigma_L)$ compressive. Correspondingly, when wood is loaded across the grain in the transverse direction, collapsing the honeycomb in Fig. 14a,

$$E_T = 0.4\, E_{S(T)} \left(\frac{\rho}{\rho_s}\right)^n \tag{5}$$

Table 1 Component compositions for dry wood

	Cellulose %	Hemicellulose (%)[a]		Lignin (%)	Other polysaccharides (%)[b]
		Glucomannan	Xylan		
Hardwood	38–51	1–4	14–30	21–31	2–4
Softwood (coniferous wood)	33–42	14–20	5–11	27–32	3–9
Cell wall layer					
Primary	15	15		70	–
S1 layer	28	27		45	–
S2 layer	45	35		20	–
S3 layer	47	38		15	–

Data from Sjöström (1993) and Bodig and Jayne (1982)
[a]See Fig. 11b
[b]Includes pectin and proteins such as starch

Table 2 Mechanical properties for wood components

Component	Density (g/cm^3)	Young's modulus E(GPa)	Tensile strength σ_y (GPa)
Cellulose (axial)	1.45–1.59	120–167	0.75–1.1
Hemicelluloses		5–8	0.1–0.2
Lignin	1.2–1.3	2.5–3.7	0.025–0.075

Data from Gibson (2012)

$$\sigma_T \cong 0.14\sigma_{S(T)}\left(\frac{\rho}{\rho_S}\right)^n, \tag{6}$$

where n can vary between 2 and 3. Normally, for a density $\rho_s \cong 1.5$ g/cm^3, longitudinal (or coaxial) Young's modulus $E_{S(L)}$ in Eq. 3 will be ~30 GPa, while in the transverse direction (across the grain), $E_{S(T)} \cong 10$ GPa in Eq. 5. The tensile strength in the longitudinal (axial) grain direction is ~0.35 GPa while it is ~0.14 GPa in the transverse direction (across the grain). Consequently, we can approximate values for $\sigma_{S(L)} \cong 0.12$ GPa and $\sigma_{S(T)} \cong 0.05$ GPa. Since the relative density for wood ρ/ρ_s can vary from 0.05 for balsa to ~0.88 for lignum vitae, Young's modulus and strength can vary by a factor of 10^3 and 10^2, respectively, while the modulus and the tensile strength of the cell wall in the longitudinal (axial) direction along the grains are up to 4.5 times those in the transverse direction, across the grains.

There are of course plants composed of fibrous, honeycomb-like cells and parenchyma closed-cell foam cells illustrated ideally in Fig. 14b where Young's modulus can be approximated by Eq. 1. For $E_s \cong 0.5$ GPa and $\rho/\rho_s \cong 0.01$, the cell modulus is given by $E \cong 0.005$ GPa. For plants like the palm or those composed of honeycomb and parenchyma foam cells, the relative density (ρ/ρ_s) can vary from ~0.03 to 0.67 and E will vary from ~0.01 to 30 GPa and corresponding strengths (σ) will range from ~0.003 to 0.3 GPa. Parenchyma cells in plants as represented schematically in Fig. 14b have a single cell wall structure consisting of cellulose, hemicellulose, and pectin.

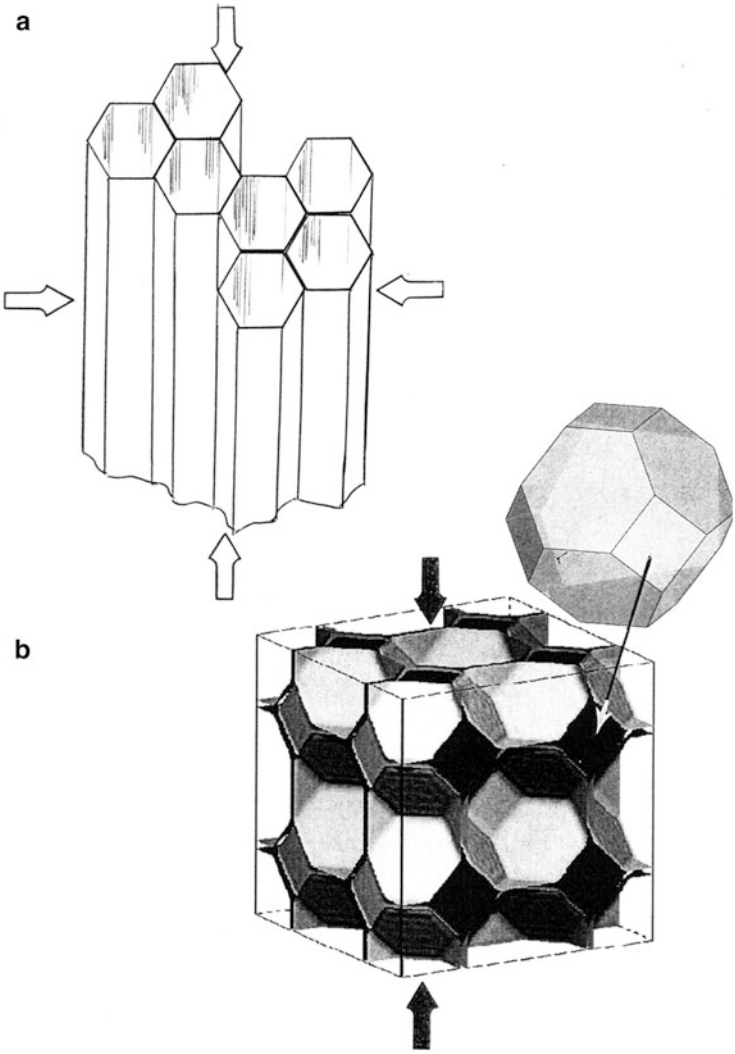


Fig. 14 (**a**) Honeycomb model for loading wood cells. (**b**) Closed (polyhedral) cell model for plant cell loading. The polyhedron, a truncated cuboctahedron, is a completely space-filling polyhedron

Examples of Wood and Plant Materials Applications

Aside from the many uses of bulk wood cut from timber along the grain structure for optimal strength, there are many applications of wood laminates and wood plies for structural beams and many varieties of plywoods corresponding to the ideal plies shown in Figs. 1g and 2a–c. Figure 15 illustrates such plywood structures based on thin sections of oriented and cross-oriented wood grains. Such building

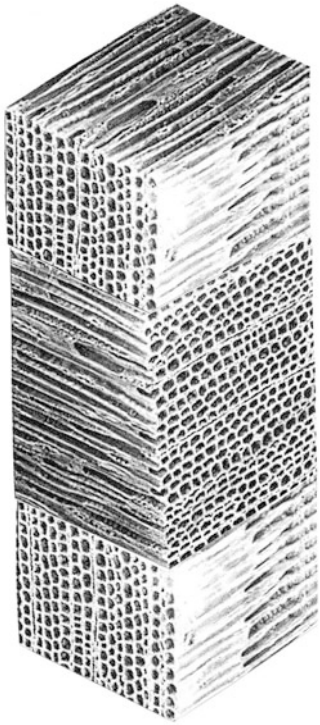


Fig. 15 Conceptual arrays of wood laminates and cross-ply structure

materials rival concrete and stone for small dwellings and structural systems. The other major application of wood especially soft woods having reduced lignin content and fast-growing cycles is pulp production for paper manufacture, although using wood to make paper was developed only around 1840. Paper made from plant pulp consisting mainly of cellulose and lignin in a pulp dates to around 200 B.C. in China and considered one of the four great Chinese inventions: the compass, gunpowder, papermaking, and woodblock printing which was first used on fabrics (textiles) prior to the invention of paper. The original Chinese paper used mulberry, hemp waste, and other bast fibers in a pulp which was drained of water to create a mat of randomly interwoven cellulose and lignin fibers which was pressed and dried into paper sheets. Even today, high-quality papers are made from pulp of cotton and other textiles, including hemp and flax. Papyrus, the Egyptian precursor of paper, also utilized a cross weave of thin strips cut from the papyrus plant stem softened in water, pounded flat, and dried in the sun.

Wood pulp is a dry, fibrous material made by mechanically or chemically separating the fibers composing wood. This pulp suspension in water can be laid down on a screen to form a sheet of paper. Wood pulp contains the wood fiber composition of cellulose, hemicellulose, and lignin (Table 1), and pulping breaks down the bulk structure into these constituent fibers while degrading the lignin and hemicellulose into small, water-soluble molecules which can be washed away from

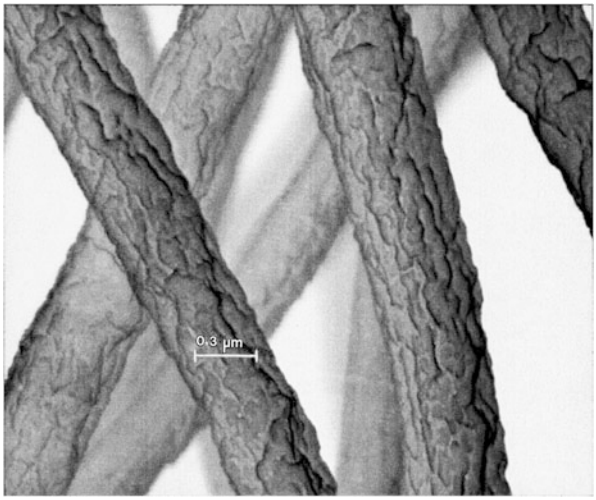


Fig. 16 Electrospun polymer-cellulose composite fiber: polymethyl methacrylate (PMMA) with 17 volume percent cellulose nanocrystals

the cellulose without depolymerizing it. As the lignin is removed, the paper product becomes softer. Consequently, lignin reduction has been achieved by gene splicing of tree tissue, a technique also used in the reverse sense to strengthen plants like barley for beer production in areas where wind and rain beat down the less rigid stalks which cannot then be harvested.

Because paper is a primarily interlaminated cellulose fiber, it can have strengths for most commercial paper ranging from 0.03 to a maximum of 0.1 GPa for high-strength paper with corresponding elastic moduli ranging from ∼2 to 4 GPa. However, processing pulp to extract smaller, nanocrystalline cellulose microfibers has allowed paper to be produced having a strength of ∼0.2 GPa, twice the highest strength of any commercial paper (Berlund 2005; Henriksson et al. 2008).

Recent efforts by the US Forest Service to produce cellulose nanocrystals from wood by-products such as wood chips and sawdust pulp have achieved yields of 30 % cellulose nanocrystals. These have been added to polymers such a polymethyl methacrylate (PMMA) (up to 17 wt%) to produce electrospun fibers illustrated in Fig. 16. These composite fibers can have strengths nearly 100 times greater than PMMA fibers (3.6 GPa in contrast to ∼0.05 GPa for PMMA) and Young's modulus similar to that of Kevlar fibers (∼131 GPa).

Of course there are a variety of other cellulosic fibers from plant materials which have been used for textile production for many millennia. Cotton, a cellular fiber having an elementary filament structure similar to wood grain illustrated in Fig. 9a, has a primary cellulose layer ∼0.1 um thick and a complex secondary layer ∼4 µm thick composed of several layers. The central filament core is referred to as the lumen and contains the cell nucleus. A wax cuticle coats the cotton fiber which resembles a thin ribbon as illustrated in Fig. 17a. Flax stem or linen fibers have a similar structure but also contain some hemicellulose and lignin, which increases

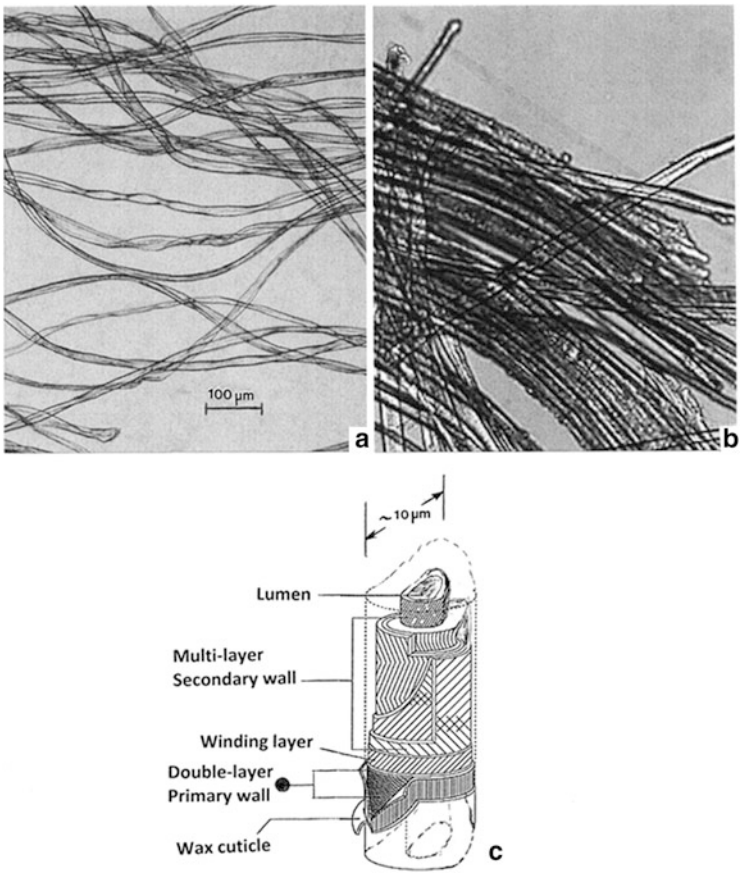


Fig. 17 Plant fibers. (**a**) Cotton. (**b**) Linen fibers composing the Shroud of Turin (Courtesy of Rachel Freer, Arizona State Museum). (**c**) Cotton microfiber structure showing multiple cellulose fiber wrapping layers. Linen microfibril structure is similar except for some hemicellulose and lignin inclusion (Table 2). Magnification of (**b**) is the same as shown in (**a**)

for other plant fibers as illustrated in Table 3, which also compares plant fiber elastic moduli and yield strengths. Figure 17b shows linen fibers while Fig. 17c illustrates the microfibril structure for cotton.

Textiles made from cotton and linen date to more than 10 millennia B.C. in Egypt and evidence of cotton in Mexico and other areas of South America date to 7 millennia B.C. Figure 17b illustrates some historical linen fibers.

It might be noted that natural fibers as illustrated for cellulose in Fig. 10 are composed of elementary fibrils forming microfibrils ~10 to 25 nm in diameter. Wood microfibrils usually average ~10 nm. As implicit in Fig. 10, microfibrils are cross-linked or otherwise formed into microfibers having diameters ranging from 0.1 to 1 μm. These bundle into fibers with diameters which can range from ~5 to 25 μm, with aspect ratios ranging from 100 to 400 um.

Table 3 Plant fiber compositions and mechanical properties

Fiber type	Density (g/cm^3)	Composition (%)					Young's modulus (GPa)	Strength (GPa)
		Cellulose	Hemicellulose	Lignin	Pectin	Other		
Cotton	1.54	>95	–	–	–	<5	10–30	0.09–0.1
Flax/linen	1.4–1.5	92	2	4	–	2	50–70	0.5–0.9
Hemp	1.41	78	10	7	3	2	30–60	0.3–0.8
Jute	1.3–1.5	65	22	11	–	2	20–55	0.2–0.5

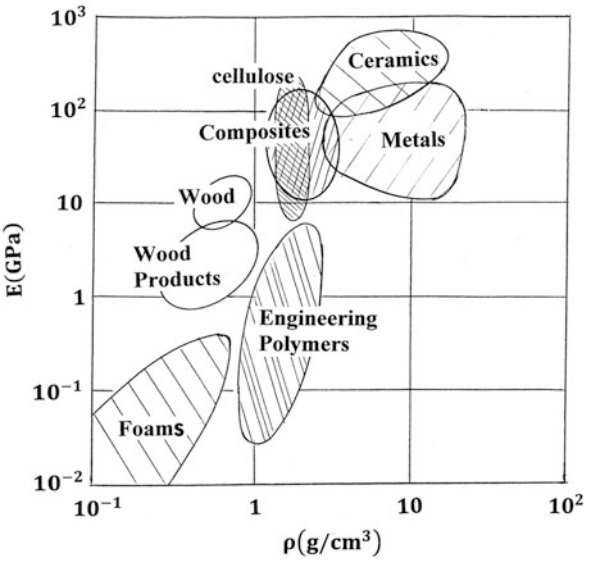


Fig. 18 Plot of Young's modulus versus density for a range of materials and composite systems

Figure 18 provides a comparative summary for cellulose and related composite moduli along with general composites, metals, ceramics, and polymer materials. Robert Moon and colleagues (2011) have written a review of cellulose nanomaterials which also summarizes these issues.

References

Adler E (1977) Lignin chemistry – past, present and future. Wood Sci Technol 11:69–218

Berlund LA (2005) Cellulose-based nanocomposites. In: Mohanty AK, Missa M, Dizal LT (eds) Natural fibers, biopolymers and their biocomposites. CRC Press, Boca Raton

Bodig J, Jayne BA (1982) Mechanics of wood and wood composites. Van Nostrand Reinhold, New York

Brunow G, Kilpelainen I, Sipila J, Syrjanen K, Karhunen P, Setala H, Rummakko P (1998) Oxidative coupling of phenols and the biosynthesis of lignin. In: Lewis NG, Sarkaner S (eds) Lignin and lignan biosynthesis, ACS symposium series 697. American Chemical Society, Washington, DC, pp 131–147

Care ID (1969) Longitudinal Young's modulus of *Pinus Radiata*. Wood Sci Technol 3:40–48

Chakraburty A, Sain M, Korstschot M (2006) Cellulose microfibers as reinforcing agents for structural materials. In: Oksmon K et al (eds) Cellulose monocomposites, ACS symposium series. American Chemical Society, Washington, DC, pp 169–186

Currey JD (1964) Three analogies to explain the mechanical properties of bone. Biorheology 2:1–10

Erb RM, Libanori R, Rothtuchs N, Studart RA (2012) Composite reinforced in 3D using low magnetic fields. Science 335:199–204

Ferguson VL, Boyde A, Bushby AJ (2005) Elastic modulus of dental enamel: effect of prism orientation and mineral content. MRS symposium proceedings 844:3–8

Fratz P (ed) (2008) Collagen: structure & mechanics. Springer, New York

Gibson LJ (2012) The hierarchical structure and mechanics of plant materials. J Roy Soc Interface 9:2749–2760
Gibson LJ, Ashby MF (1982) The mechanics of three-dimensional cellular materials. Proc Roy Soc London A Math Phys Sci 383:43–59
Gibson LJ, Ashby MF, Harley BA (2010) Cellular materials in nature and medicine. Cambridge University Press, Cambridge
Glimcher M (2006) Bone. Nature of the calcium phosphate crystals and cellular, structural, and physical chemical mechanisms in their formation. In: Schoonen SH (ed) Medical mineralogy and geochemistry. Reu Mineral Geochem 64:223–282
Henriksson M, Berglund LA, Isaksson P, Lindstrom T, Nishino T (2008) Cellulose nanopaper structure of high toughness. Biomacromolecules 9:1579–1585
Kirkham J, Brookes SJ, Shore RC, Bonass WA, Smith DA, Wallwork ML, Robinson C (1998) Atomic force microscopy studies of crystal surface topology during enamel development. Connect Tissue Res 38:91–100
Lowernstam HA, Weiner S (1989) On biomineralization. Oxford University Press, New York
Mack RW (1964) Internal publication. University of California, San Francisco. In reference [225]
McKee MD, Addison WH, Kaartinen MT (2005) Hierarchies of extracellular matrix and mineral organization in bone of the craniofacial complex and skeleton. Cells Tissues Organs 181:176–188
Moon RJ, Martini A, Maira J, Simonsen J, Youngblood J (2011) Cellulose nanomaterials review: structure, properties and nanocomposites. Chem Soc Rev 40:3941–3994
Murr LE, Ramirez DA (2012) The microstructure of the cultured freshwater pearl. JOM 64 (4):468–474
Piekarski K (1973) Analysis of bone as a composite material. Int J Eng Sci 11:557–565
Povolo F, Hermida EB (2000) Measurement of the elastic modulus of dentin pieces. J Alloys Compound 310:392–395
Sjöström E (1993) Wood chemistry. Fundamentals and applications, 2nd edn. Academic, San Diego, p 293
Staines M, Robinson WH, Hood JAA (1981) Spherical induction of tooth enamel. J Mater Sci 16 (9):2551–2556

Examples of Man-Made Composite Structures

Contents

Abstract

Some examples of man-made composites are presented in this chapter, sporting goods, wind turbine blades, aircraft structures, automobile structures, and home construction, which can provide an example of a broad range of composite structures in everyday life.

Introduction

There are almost as many man-made composites or composite systems and structures utilized in construction and manufacturing as there are in the natural plant and animal worlds. One of the principal materials of construction is of course wood and wood laminates of various kinds, as well as the use of woven natural fibers including cotton, silk, linen, hemp, and even cellulose or composite cellulose fibers as illustrated in ► Fig. 16 in chapter "Examples of Natural Composites and Composite Structures."

Among the more notable man-made fibers and fiber commodities are carbon fiber weaves and aramid or Kevlar (poly-*p*-Phenylene-terephthalate). Kevlar primary fibrils are composed of hydrogen-bond cross-linking (Fig. 1a) similar to cellulose (► Fig. 11a of chapter "Examples of Natural Composites and Composite Structures"). These form multifilament fibers as shown by the arrow in Fig. 1b. Strong carbon or graphite fibers have a Young's modulus near 450 GPa, half the modulus of multiwall carbon nanotubes (Kashyap and Patil 2008), while Kevlar

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_26

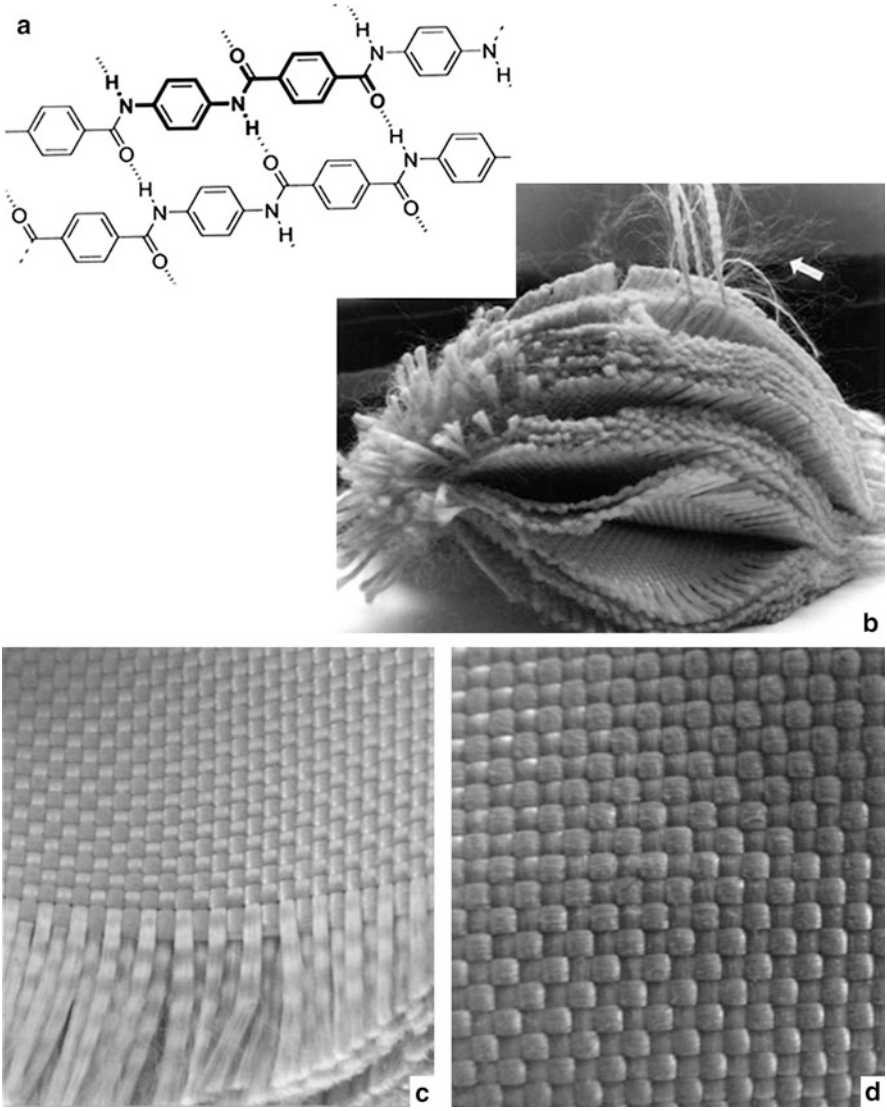


Fig. 1 Kevlar fabric. (**a**) Molecular structure of Kevlar. *Bold* component of the molecule represents the monomer structure. *Dotted lines* are hydrogen bonds. (**b**) 21–90° weave fabric layers composing a lightweight bulletproof vest. *Arrow* shows fibril strand composing multi-stand fibers. (**c**) and (**d**) show enlarged view of the weave

fiber can vary from ~90 to 200 GPa, with strengths of 3.6 GPa. As standard woven fabric, Kevlar is best known for body armor applications where bulletproof vests can contain from 20 to 60 overlain woven Kevlar layers as illustrated in Fig. 1. In many composite applications of more rigid design, such as helmets, these layers (Fig. 1b and d) are impregnated with a suitable matrix such as epoxy and glued

together. Impregnated fiber weaves are referred to as prepregs and are fabricated so that the epoxy matrix is not completely cured, often by storing in a cold environment. These prepregs are more flexible and can be molded or wrapped using additional epoxy or binder to create a monolithic composite with layers of woven fiber fabrics bound in a continuous polymer matrix. The fabric weave or pattern as illustrated ideally in ▶ Fig. 2 (chapter "Classification of Composite Materials and Structures") can be arranged to form an effective laminate structure to optimize mechanical performance relative to the imposed stresses or strains, impact, fatigue, or related mechanical circumstances. These can be overlain on formed cores or in complex composite structures including foam layers or honeycomb sandwiches (▶ Fig. 14 in chapter "Classification of Composite Materials and Structures"), including wood laminates or veneers.

Composite Sporting Goods

Some of the most common sporting goods – baseball bats, rackets, skis, canoes, boats, etc. – were originally constructed of wood, and many items continue to be made from wood or wood laminates. Baseball bats have been traditionally fabricated as solid wood oriented along the wood grain and are still used in professional (Major League) baseball in the United States. While having requisite weight and strength (stiffness), wood bats can break, and moisture variations can alter the optimized properties of the bat. Recently, extruded aluminum bats have been fabricated and used extensively for baseball and softball games, but not in professional play. In the past decade, there have been numerous composite baseball bat configurations including carbon/epoxy or Kevlar/epoxy or uniaxial wraps around an extruded aluminum core and with wood surface veneer, as well as similar composite weave or unidirectional wrappings around tapered wood cores illustrated conceptually in Fig. 2a.

Racket handles and frames for tennis and badminton present similar composite opportunities, and modern tennis rackets (Fig. 2c) are popularly fabricated from carbon–fiber/epoxy and related composites as wraps upon light metal frames such as titanium and titanium alloys. Similar structures are utilized in fabricating golf club shafts. Racket frames and golf club shafts are also fabricated from multiwall carbon nanotube composites (Khare and Bose 2005; Ci et al. 2008). Tennis and badminton racket strings, as illustrated in Fig. 2c, are often nylon or polyamide fibers.

Figure 2b illustrates a bicycle portion illustrating novel use of carbon fiber/epoxy composite to create a strong, open, and correspondingly lightweight structure. Even the wheels and spokes for bicycles are fabricated from carbon composites, especially high-strength carbon fiber composites. Figure 2 illustrates some rather similar and typical design and fabrication strategies utilizing unidirectional or laminate weave and oriented weave composites (▶ Fig. 2 of chapter "Classification of Composite Materials and Structures"), often referred to as unidirectional layups (lamina) or quasi-isotropic layup laminates.

Unidirectional hybrid, glass/carbon epoxy composites are also used in the manufacture of fishing rods. Similar composite structures are used in fabricating

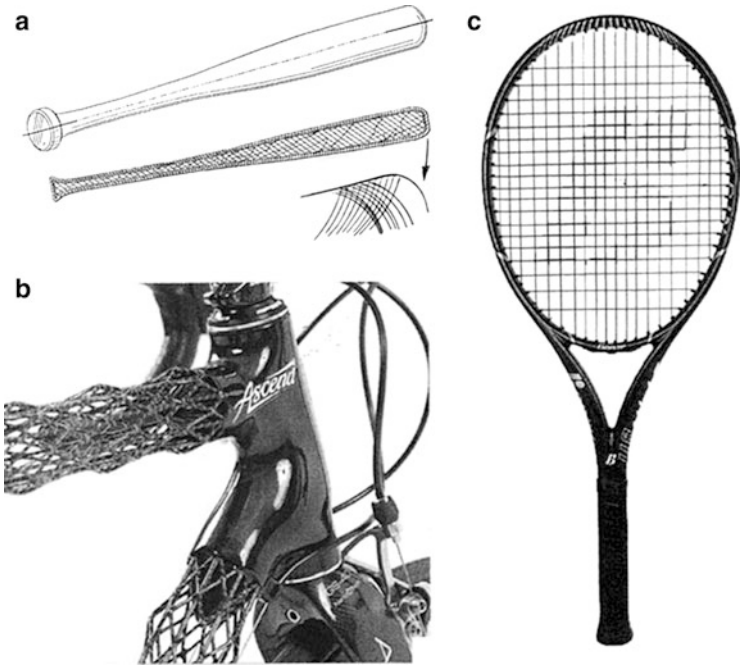


Fig. 2 Sports composites and composite concepts. (**a**) Baseball bat showing overlay of cross-ply of directional fibers in matrix such as epoxy on aluminum or wood core. (**b**) Lightweight, open structure carbon fiber and rod construction for bicycle frame. (**c**) Tennis racket fabricated from carbon fiber/epoxy composite

other, similar long, flexural systems. Figure 3 illustrates the incorporation of laminate or multi-laminate composites with foam sandwich or honeycomb designs for snow skis, water skis, and snowboards. In these structures, light weight, high strength and stiffness, flexibility, and tension strength, especially in high performance skis, require novel composite construction which often incorporates wood or wood laminates as well (Froes and Haake 2001; Jenkins 2003). As illustrated in Fig. 3a, modern snow ski and board fabrication are characterized by a top sheet, a core, and a base, with metal (steel) edges. Base materials are often high molecular weight polyethylene which can be coated with a layer of wax on the bottom (or exposed) side. Edges are made of steel or stainless steel. Ski cores are often laminated hardwoods such as birch, aspen, ash, beech, etc., generally with different woods laminated together. Some skis have incorporated titanium, while polyurethane foams sandwiched between carbon on Kevlar fiber laminate with an adhesive layer, or aluminum or laminate honeycomb with sandwich or facing sheets (skins) of carbon or Kevlar fabric in epoxy, with an adhesive layer between the honeycomb and the facing sheet. There are often laminates above and below the core structure and rubber damping strips between the composite layers as well as a sidewall configuration on either side of the core for ski construction. In snowboard fabrication, the core is larger (thicker) and assumes a more prominent design

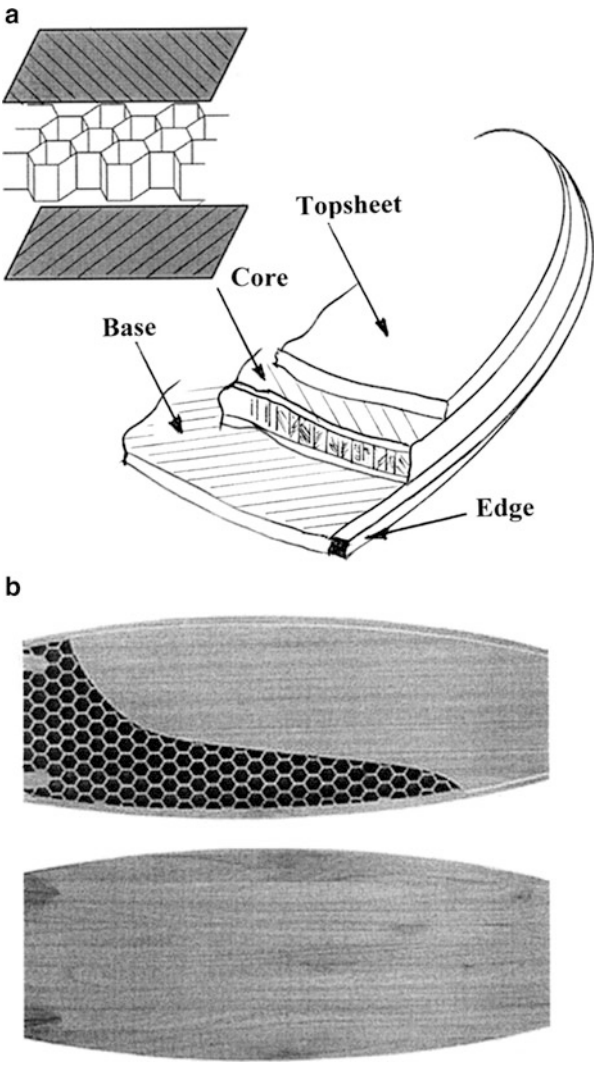


Fig. 3 Ski and snowboard composite structure (**a**) and wood/aluminum or other lightweight, high-strength honeycomb surfboard structure (**b**). In (**a**), the basic honeycomb core structure consists of a hexagonal, high-strength honeycomb sandwiched between an adhesive layer and a facing sheet or skin. Several laminate composite layers can be included above and below the core regime

consideration. Foams are less prominent in snowboard cores, and carbon fiber rods are often added. Wood laminates also continue to be used, and as illustrated in Fig. 3b, wood construction in surfboards using honeycomb cores continues to be popular since it contributes to light weight and requisite mechanical performance.

Fiberglass/epoxy regimes (laminates) above and below the laminated core, especially hardwood cores, in snowboards is also a popular composite fabrication strategy because of its lower cost in contrast to carbon fiber composites. These fiberglass layers also add increased stiffness to snowboard weave or wrap design (with fibers woven together at 90°) and triaxial wraps where the fiberglass strands are woven at +45°, 0°, and -45°, which provides increased torsional stiffness.

Fig. 4 Wind turbine with 75 m long blades

The composite fabrication strategies implicit in sporting goods manufacture as illustrated in Figs. 2 and 3 are also utilized in larger-scale production arenas such as airplane and other aerospace manufacture. Turbine blade design and manufacture have also incorporated composites and composite laminate technologies.

Wind Turbine and Aircraft Composite Structures

Wind turbine blades illustrate the integration of many composite systems in the fabrication of larger-scale composite structures. Standard 1.5 MW (megawatt) wind turbine blades measure 35–40 m in length and weigh 6–7 t. In 2012, the world's largest blade manufactured by Siemens Energy in Hamburg, Germany, for a 6 MW offshore wind turbine measured 75 m in length and is illustrated in Fig. 4. Figure 5 shows a schematic cutaway view of the complete wind turbine components.

A typical wind turbine blade construction consists of outer skins supported by a main spar. The outer skin construction consists mainly by laminates or a sandwich shell which is often either a foam or honeycomb with a film adhesive bonding a high-strength composite laminate facing on either side of the sandwich. Blade skins often employ glass (or fiber glass)/epoxy (e-glass) while longer blade construction. Figure 4 utilizes high-strength carbon fiber/epoxy laminates. As shown in Fig. 6, turbine blade manufacturing involves four key components: the root section, the structural spar or spar cap, the aerodynamic fairing or shell, and the surfacing which consists of an in-mold primer followed by a subsequent painting operation. The blade shells typically contain triaxial and biaxial e-glass fiber/epoxy laminates using prepregs or similar manufacturing strategies. The spar or main-bearing structure uses unidirectional fiber construction of either glass or carbon fibers, with laminate sections up to 5 cm thick toward the root section.

The root section of the wind turbine blade uses metal inserts bonded into the composite using adhesive or infusion resins where the composite thickness

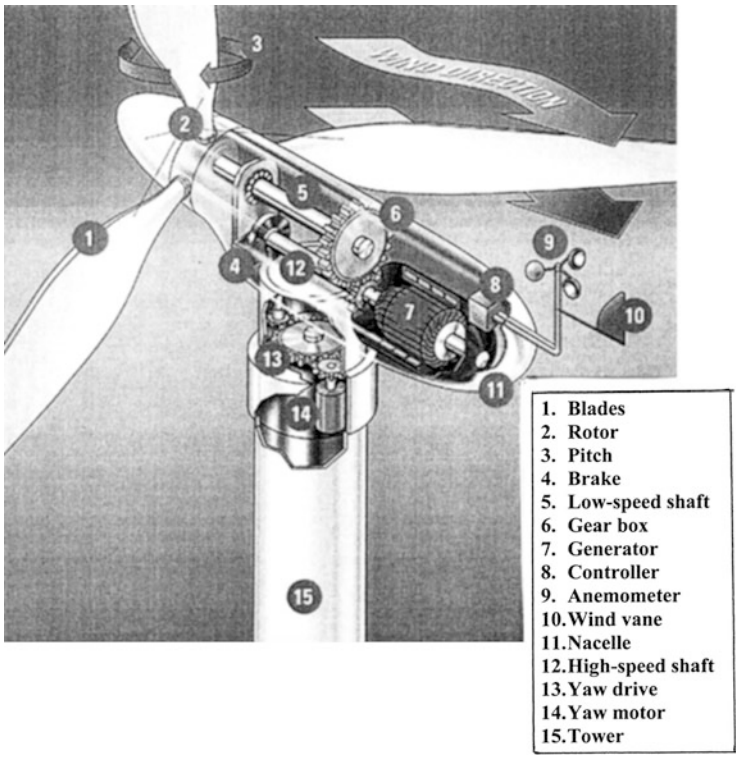


Fig. 5 Wind turbine cutaway showing components

must be 5–10 cm thick depending on the root design. The root section fixes the blade to the rotor (Fig. 5) and transfers the loads from the composite structure to the hub and main drive shaft. Blade design must optimize low weight and rotational inertia, rigidity, and resistance to fatigue and wear over a roughly 20 year service life. Wear features include UV degradation, rain, hail, and salt aerosol in offshore units.

Composite honeycomb structures have been used in numerous aerospace products and components including helicopter blades and components, reentry vehicle skins and structures, as well as a variety of aircraft component structures utilizing a variety of high-strength fiber fabrics and fabric or prepreg laminates, especially high-strength carbon (graphite)/epoxy composites. Typical fighter aircraft composite use has increased over the past two decades along with commercial airplane composite structures which currently exceed 50 % of the total materials for construction, especially the airframe. Figure 7 illustrates the range of composite materials in the main structure of a Boeing 757–200 commercial aircraft. Composite structure employing carbon fiber arrays are particularly attractive since they significantly reduce the weight and increase the strength in contrast to both aluminum and titanium or titanium alloys such as Ti–6Al–4V.

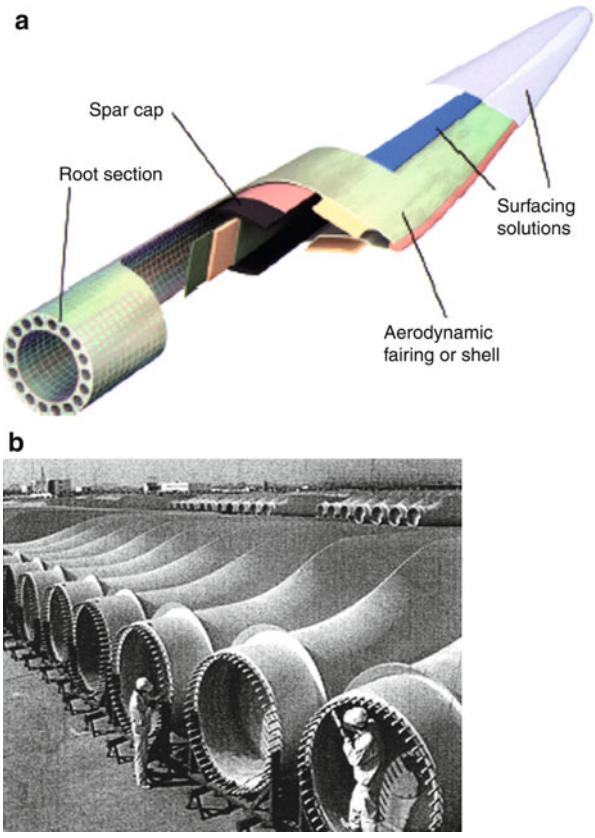


Fig. 6 Wind turbine blade schematic (**a**) and manufactured blades showing root sections (**b**)

Structural Composites in Automobiles

Concerns for lightweight materials in cars were an issue in early manufacturing, and aluminum and magnesium were used in automobile manufacturing as early as 1936. Both the Ford Thunderbird and Chevrolet Corvette used glass fiber-reinforced polymer (epoxy) composites in the 1950s, and the first structural composite road car was the Lotus Elite in 1957. However, during this period, the Marcus sports car used plywood for a Formula 2 chassis in contrast to most other racing cars of this period which used tubular steel frames for the primary structure.

Carbon fiber/epoxy composites and honeycomb panels were used in the automotive industry in the early 1980s and became the preferred choice for racing cars by the mid-1980s.

Since the 1990s, the automotive industry has increased the uses for fiber-reinforced composites, especially carbon/epoxy structures, although cost has been a continuing concern when considering both carbon fibers and the epoxy matrix or prepreg. Recently, high-end automobiles such as Lamborghini have incorporated

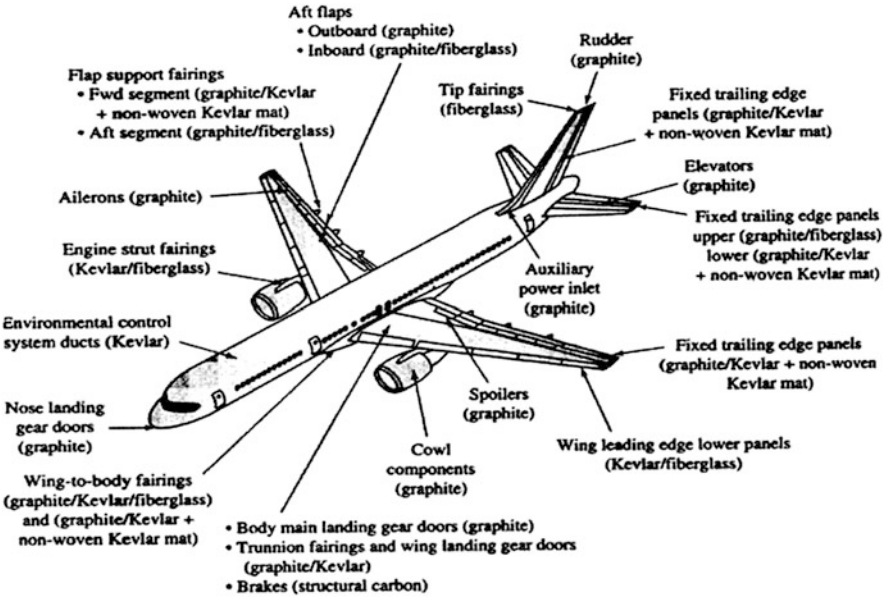


Fig. 7 Composite ports in the main structure of the Boeing 757–200 aircraft (Source : Boeing Commercial Airplane Co)

high-strength carbon fiber-reinforced composites throughout the vehicle, including structural frame members, body panels, exterior components, and the like as illustrated in Fig. 8a.

The majority of fibers in composites such as carbon fiber are utilized in woven textile formats illustrated conceptually in Fig. 1c and d for Kevlar. These woven textiles also include glass fiber textiles, aramid (Kevlar), and even natural fibers in some cases. Low-cost matrix polymers are also used in some applications, and even recycled paper has been used in honeycomb core construction for some automotive panels. Primary structures in many commercial automobiles (Fig. 8b) continue to employ steel, including the frame and chassis, but smaller hybrids and electric cars are using carbon fiber composites in structural components as illustrated in Fig. 8c. While unidirectional carbon fibers compose some structural panels, newer applications are employing flat yarn weaves in a ribbonlike structure and with so-called five-harness or similar weave architectures as illustrated in ▶ Fig. 2e of chapter "Classification of Composite Materials and Structures." These are used in wraps and layups for panels and exterior structures such as roof and door panels (Fig. 8c).

One of the most common composite items on automobiles and other vehicles involves the tire. Tires are composed of vulcanized rubber mixed with around 30 % (weight) black carbon nanoparticles resembling the high-resolution TEM image shown in ▶ Fig. 20 in chapter "Chemical Forces: Nanoparticles" and the soot particles shown in Fig. 22. The black carbon reinforces the rubber, and as the tire wears, it is released into the environment in essentially the same form as it was

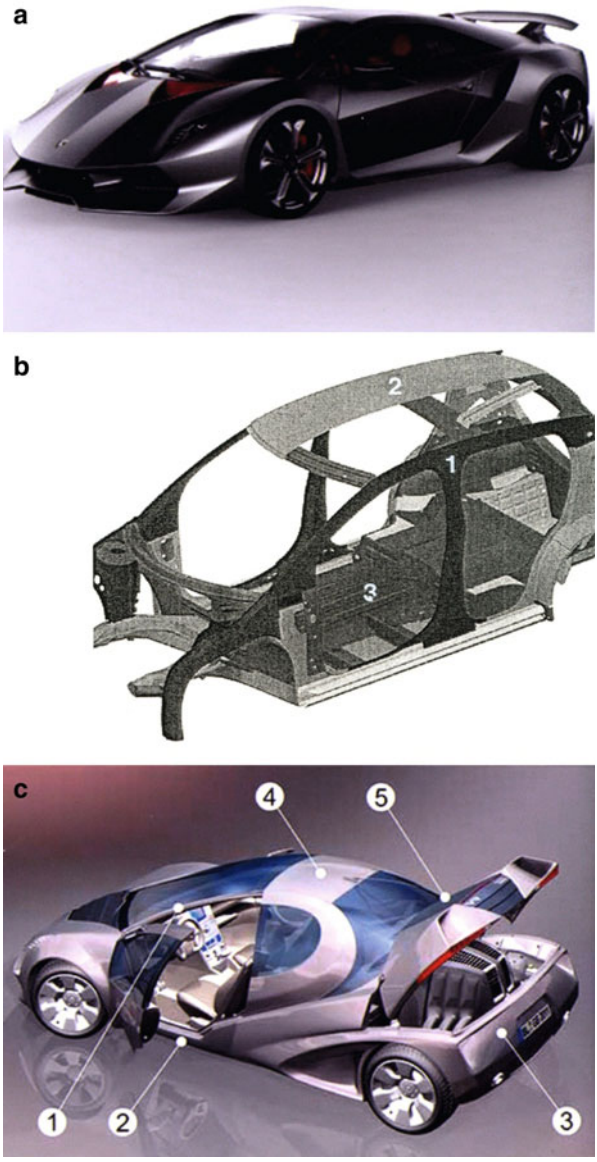


Fig. 8 Composites in automotive structures. (**a**) 2010 carbon-composite Lamborghini. (**b**) Automotive frame structures. (*1*) Frame unidirectional and multi-ply composites. (*2*) Roof laminates. (*3*) Underfloor laminates and honeycomb. (**c**) Carbon fiber composite automobile uses. (*1*) In vehicle frame, (*2*) underfloor reinforcement, (*3*) rear bumper, (*4*) roof construction, and (*5*) bonnet (Courtesy of Groz-Beckert)

originally mixed with the rubber. The construction of a tire also utilizes nylon fiber or Kevlar fiber weaves and so-called belts, which are also steel or steel-rubber laminates in many cases or layers of fiber lamina. Figure 9 illustrates schematically general tire composite construction along with the molecular structures for natural rubber (Fig. 9b), sulfur-vulcanized rubber (Fig. 9c), and nylon (Fig. 9d). In vulcanization, sulfur cross-links the rubber molecules forming a more stable and heat-resistant polymer. Styrene–butadiene rubber, used for most passenger cars, is

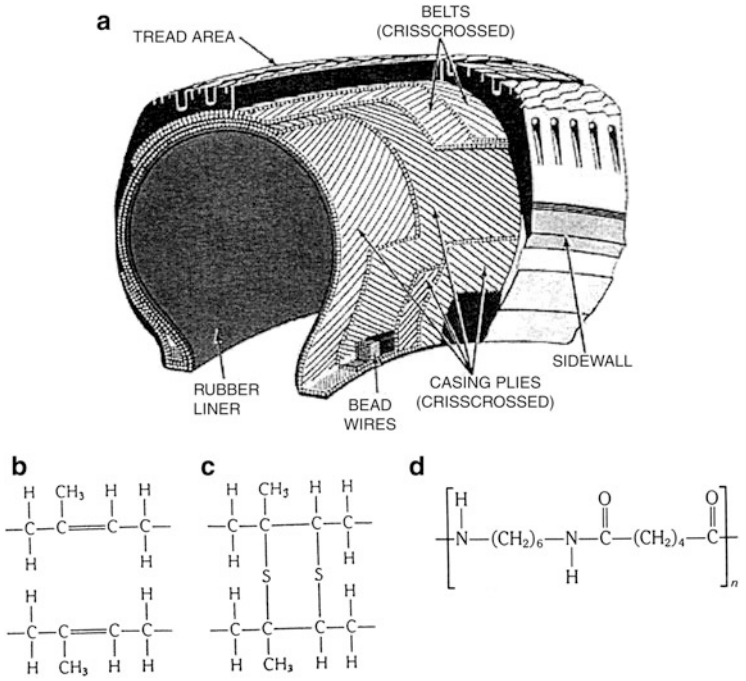


Fig. 9 Composite tire construction. (**a**) Tire construction/component schematic. (**b**) Natural rubber. (**c**) Vulcanized (cross-linked) rubber. (**d**) Nylon structure

also vulcanized as shown in Fig. 9c. The beads (Fig. 9a) or bands of wires are brass (Cu–Zn)-coated high-strength steel encased in a rubber compound to provide strength to fit the tire to the wheel.

Energy-absorbing polymeric forms and laminated honeycomb structures are also widely used in the automotive industry to prevent front or side collision injuries. These structures absorb kinetic energy by compressing or deflecting at a relatively constant stress during the period which is defined as a crash plateau as illustrated in the ideal foam stress–strain diagram shown in Fig. 10. This plateau can extend to roughly 50–20 % of strain, and the absorbed energy can be represented by the area under the crush plateau:

$$W = \int_{o}^{\in} \sigma \cdot \in d\in \tag{1}$$

As shown in Fig. 10, when the impact load is applied to a foam structure, it will yield elastically up to a strain of ~5 %, and after this strain, it will begin to buckle and collapse continuously at relatively constant strain, absorbing energy of impact. The point in the stress–strain diagram of Fig. 10 where plastic deformation commences defines a so-called crash strength of the foam, σ_{cr}. Generally, for a foam material,

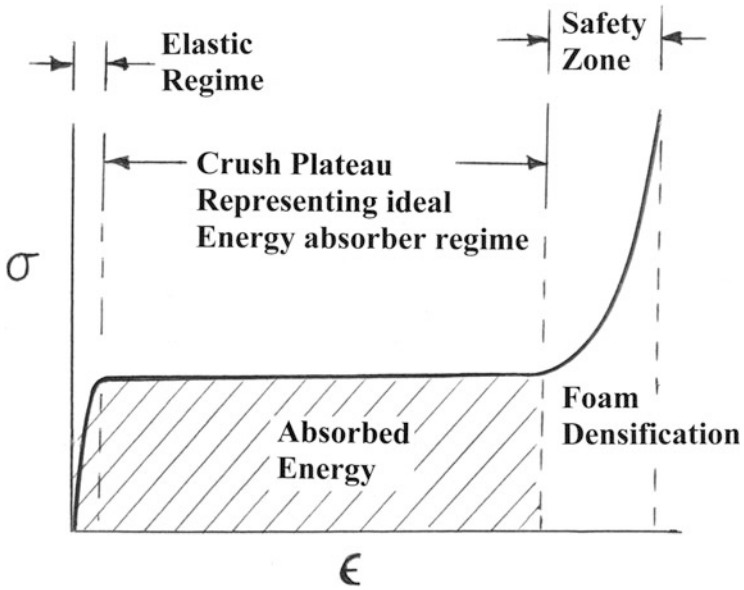


Fig. 10 Typical (ideal) foam stress–strain diagram. The absorbed energy can include the elastic stored energy which is relatively small in comparison to the crush plateau. Actual curves are usually very irregular because of irregular crushing

$$\sigma_{cr} = 0.58\sigma_y(\rho/\rho_s)^{3/2} \tag{2}$$

where σ_y is the tensile yield strength of the solid material of the foam having a density ρ and ρ_s is the solid density. The coefficient 0.58 in Eq. 2 is based on the evaluation of a wide range of foam materials, both polymeric and metallic, and largely applicable to closed-cell foams illustrated ideally in ▶ Fig. 14b of chapter "Examples of Natural Composites and Composite Structures."

The crushing force (p_{cr}) for axially crushed hexagonal (bare) honeycombs has been expressed by Alghambi (2001) as

$$p_{cr} = 8.6\sigma_y t^{5/3} C^{1/3} \tag{3}$$

where ρ_{cr} is the mean crushing force, σ_y is the tensile yield strength for the solid honeycomb material, t is the honeycomb cell thickness, and C is the minimum width of the honeycomb (side width). Correspondingly, the minimum cell thickness-to-width ratio needed to prevent elastic buckling collapse during axial loading of a thin-wall square column has been given by Zuidema (2012) as

$$(t/C) \geqq 0.5\left[\sigma_y\left(1-\nu^2\right)E\right]^{1/2}, \tag{4}$$

where ν is Poisson's ratio for the solid material and E is the Young's modulus; and this equation can be generally applied to a hexagonal honeycomb. It can be observed in Eq. 4 that the Young's (elastic) modulus is a major feature for

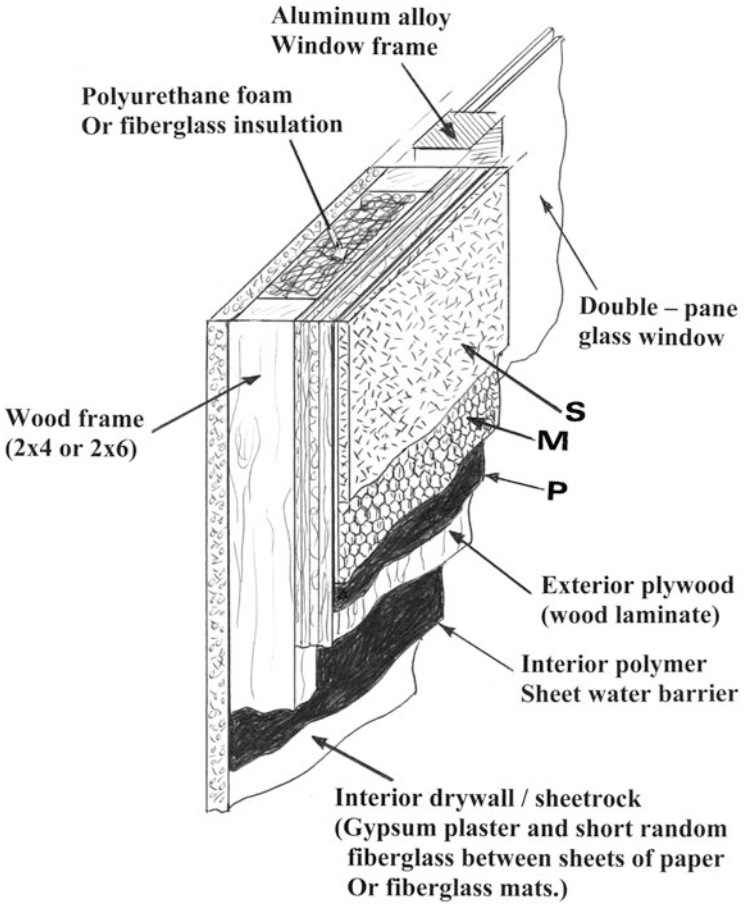


Fig. 11 Common exterior wall structure for wood-framed home representing multi-laminate, multi-composite structure. The exterior stucco (*S*) composed of Portland cement, sand, and short, random polymer fibers is placed on a hexagonal metal (*M*) (steel) mesh structure which overlays a pitch (*P*)-based water barrier usually a fiberglass, pitch impregnated mat

honeycomb energy absorption, and cores constructed from high stiffness materials such as carbon and especially uniaxial carbon fiber composites can significantly improve crash safety.

Energy-absorbing composite and sandwich aways for energy absorption are also used in ship hulls and aircraft fuselages (Fig. 7). Multi-laminate and multilayer composite arrays are also utilized in bulk armor, also employing foam sandwiches as well as thick laminates of metal/ceramic or ceramic/metal multilayers to create projectile trajectory diversions and momentum interruption. Bulletproof vests also employ an outer, thick ceramic plate, such as alumina (Al_2O_3) or TiB_2, to create trajectory diversion and momentum interruption before a high-energy projectile strikes the multilayer Kevlar fabric.

While carbon fibers, especially high-strength carbon fiber, are finding increasing uses and are replacing many high-strength metals and alloys, the use of carbon nanotubes has also increased over the past decade and will become a significant composite fiber when it can be spun into useful yarn or textile fiber systems. In addition to carbon nanotube composites, nanocomposites in a more general sense pose some interesting prospects, especially since the behavior of nanomaterials is often much different than bulk materials. We will in fact address these issues more specific to nanomaterials in a later chapter.

While structural composites in automobiles have been examined (Figs. 8 and 9), the automobile is an even more complex composite system containing metals and alloys, ceramics, electronic and magnetic materials, and a host of polymers aside from structural composite matrices. In addition, the treatment of natural composites included bones and teeth, but like the automobile, humans and other living organisms are far more complex composite systems (Fig. 11).

Additional discussion of natural, living biological, and related materials will continue in parts VIII and IX. Of particular interest will be the mimicking of biological materials and biological materials systems in the context of biomimetics.

References

Alghambi AAA (2001) Collapsible impact energy absorbers: an overview. Thin-walled Struct 39:189–213

Ci L, Suhr J, Pushparaj V, Zhang X, Ajayan PM (2008) Continuous carbon nanotube reinforced composites. Nano Left 8(9):2762–2766

Froes FH, Haake SJ (eds) (2001) Materials and science in sports. TMS, Warrendale

Jenkins M (ed) (2003) Materials in sports. Woodhead Publications, New York

Kashyap KT, Patil RG (2008) On Young's modulus of multi-walled carbon nanotubes. Bull Mater Sci 31(2):185–187

Khare R, Bose S (2005) Carbon nanotube based composites – a review. J Miner Mater Charact Eng 4(1):31–46

Zuidema BK (2012) Bridging the design-manufacturing-materials data gap: material properties for optimum design and manufacturing performance in light vehicle steel-intensive body structures. JOM 64(9):1039–1047

Part VIII

Biological Materials

Structure and Function of Viruses and Bacteria

Contents

Abstract

Proteins are the building blocks of biological materials, and fundamental to these building blocks are 20 amino acids. The evolution of RNA and DNA is also fundamental to living materials and to primary life forms such as viruses and bacteria. This chapter begins with a historical perspective involving DNA. The structures and functions of virus and bacteria are then described in relation to both RNA and DNA. The catalytic role of bacteria is illustrated in specific examples involving the leaching of copper from porphyry copper waste.

Introduction

While proteins constitute the elementary building blocks of most biological materials, their structure is a polypeptide chain consisting of a sequence of amino acid residues. These chains cross-link, fold, and twist to form larger length-scale hierarchies. The functional diversity exhibited by these biomolecules is related to the combinational possibilities of the 20 monomeric amino acid units composing them (Fig. 1). Each codon of the deoxyribonucleic acid (DNA) codes for or specifies a single amino acid, and each nucleotide unit consists of a phosphate, deoxyribose sugar, and one of the four nitrogenous nucleotide bases: adenine (A), guanine (G), cytosine (C), and thymine (T). These bases are paired together by hydrogen bonds in the antiparallel sequences combined to form the DNA double helix.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_27

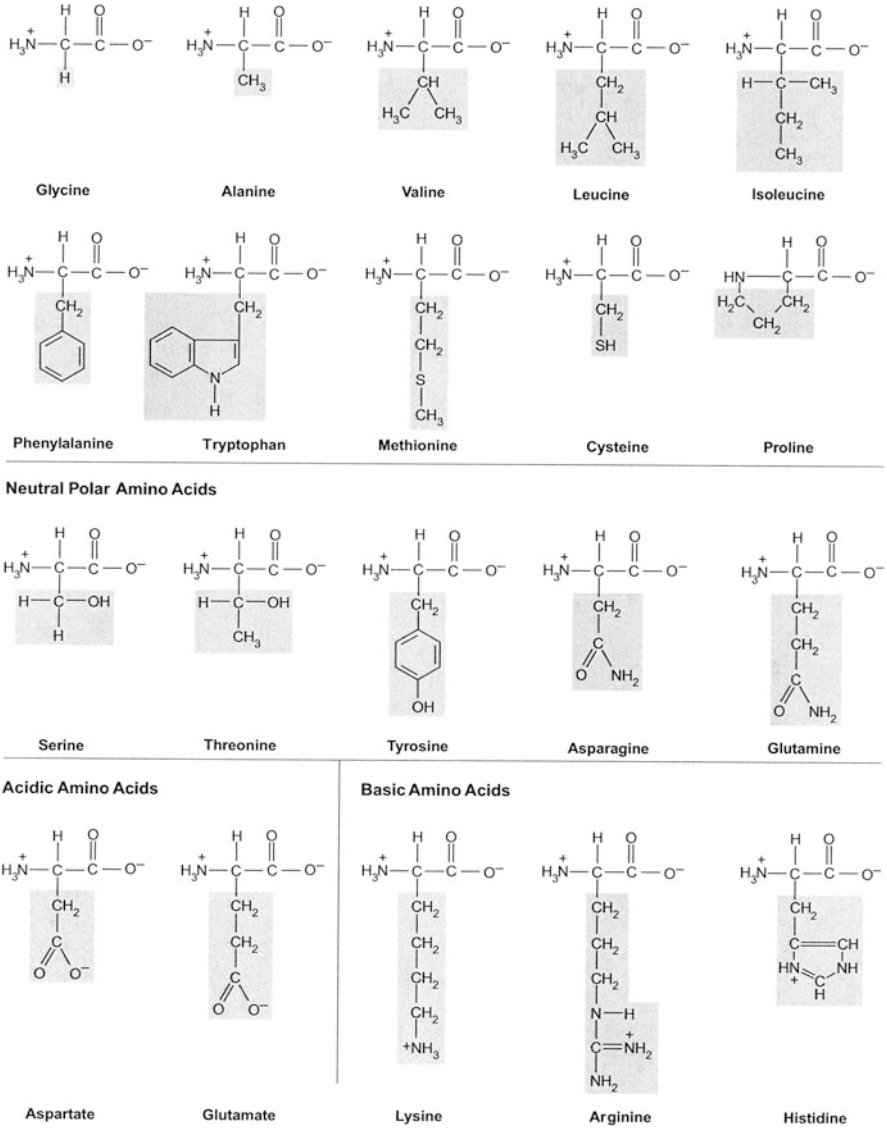


Fig. 1 The standard (20) amino acids. The *shaded boxes* represent the side chains. Those are referred to as neutral, nonpolar amino acid units

Figure 2 shows a graphic cartoon illustrating the development of the DNA double helix as a consequence of a few key chemical structural phenomena. The larger DNA "snake" evolves from the structure of benzene containing a six-membered ring of carbon atoms with alternating single and double bonds (as well as hydrogen bonded to each carbon) first described by Friedrich August Kekule in 1865. It was purported that this structure came to Kekule in a dream (the

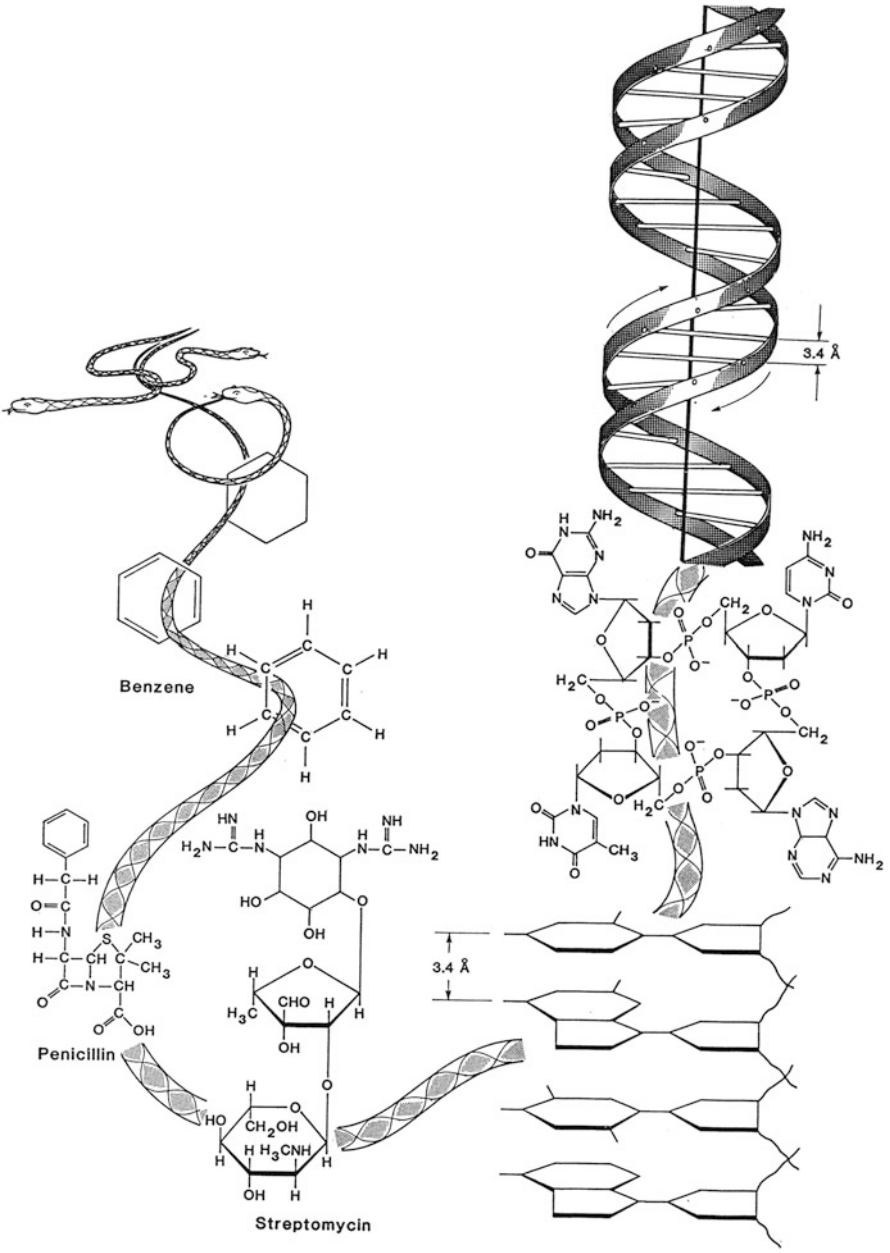


Fig. 2 Graphic cartoon drawn by the author around 1985 representing DNA double helix as the head of an evolving snake from the snake dream of Kekule which proposed the benzene ring structure forming the basis of organic chemistry. Structures of penicillin discovered by Alexander Fleming in 1928 and streptomycin by S. A. Waksman in 1943 and awarded the Nobel Prize in Physiology/Medicine in 1952 are represented in the graphic evolution of DNA. Penicillin provided the origins of antibiotics, while streptomycin was the first antibiotic effective against tuberculosis

Fig. 3 Linus Pauling (1901–1994) was one of the most influential chemists in history and one of the most important scientists of the twentieth century. Pauling is the only person to be awarded two unshared Nobel Prizes. He received the Nobel Prize in Chemistry in 1954 and the Nobel Peace Prize for 1962 as a consequence of his long campaign against nuclear weapon testing

ouroboros dream) where a snake bit its tail to create the ring structure implicit in benzene. Kekule is credited as the principal formulator of the theory of chemical structure about 10 years prior. This theory embodied the idea of atomic valence and especially the ability of carbon atoms to bond to each other and established the field of organic chemistry.

While Kekule originally proposed that his benzene molecule oscillated between two equivalent structures in such a way that the single and double bonds between carbon atoms continually changed position, a more complete theoretical basis was later proposed by Linus Pauling (circa 1928) (Fig. 3) who replaced the oscillation concept by one involving resonance between quantum mechanical structures. This concept of the resonance of molecules among several valence-bond structures became a prominent part of Pauling's monumental work, *The Nature of the Chemical Bond*, published in 1939. This seminal work was recognized with the award of the 1954 Nobel Prize.

As early as 1933, Pauling became involved in nucleic acid research when he hypothesized a structure for guanine, a base ring. In 1951 Pauling, Robert Corey, and Herman Branson proposed the α-helix and β-sheet as the primary structural forms of protein secondary structure. This included a triple helix which for Pauling provoked the hypothesis that DNA was a triple helix with large base groups facing out and the phosphate groups stacked in the core. In 1953, Pauling and colleague Robert Corey at Cal Tech published "A proposed structure for the nucleic acids" in the Proceedings of the National Academy of Science (Pauling and Corey 1953) where they stated the proposed structure was "promising." This structure, the triple helix, proved to be wrong. Only 3 months after this paper was published, Pauling visited England in April of 1953 on his way to a conference to see the James Watson and Francis Crick double helix model of DNA as represented in the "head" of the snake in Fig. 2, which he conceded was the correct model. Figure 2 illustrates the 3.4 Å distance along the molecular structure proposed by Pauling and identified in X-ray data of W. T. Astbury which is also indicted in the correct DNA structure (Fig. 2).

It is worth noting that the discovery of the so-called wonder drug, streptomycin, part of the graphic in Fig. 2, represents one of many great scientific disputes. While Selman Waksman is credited with its discovery and received a Nobel Prize recognizing that credit, Albert Schatz, a graduate student in Waksman's laboratory at Rutgers University, actually made the discovery in experiments designed to find an antibiotic against Gram-negative organisms. Waksman deliberately disenfranchised Schatz who was at least the codiscoverer of streptomycin (Schatz et al. 1944).

Proteins are amino acid polymers containing at least 50 amino acid units or residues as they are sometimes called. The standard amino acids are illustrated in Fig. 1. Molecules having molecular weights ranging from ~10^3 to 10^6 Daltons (D) are called polypeptides, while those of lower molecular weights consisting of fewer than 50 amino acid units are called peptides. Consequently, each protein consists of one or more polypeptide chains. In each protein, the types and quantities of amino acids are covalently linked in a linear (chain) sequence which is specified by the DNA base sequence generated in RNA for that specific protein. It is the uniqueness of the amino acid sequencing in the protein structure which makes it unique both in structure and function. The interaction of the amino acid residues causes subsequent folding of the molecule into a complex, 3D, and biologically active structure, dictated by the inherent information in the amino acid structures (Fig. 1). These complex protein molecule structures create the hierarchy in developing cells, organs (or organisms), and the evolution of biological systems which, as discussed by Ackbarow and Buehler (2011), illustrate the "coexistence of universality and diversity through hierarchical design – referred to as the universality – diversity paradigm as an overarching feature in protein structures."

While proteins observed in chapter "▶ Examples of Natural Composites and Composite Structures" are building blocks in the hierarchical development of cells and tissue biological structures, they also act as drugs and vaccines. In addition, it is possible to design protein structures or to produce them using both viruses and bacteria where bacteria in particular have been referred to as protein production factories or bacterial cell factories (Villaverde 2010; Vasquez and Villaverde 2010). Similarly, viruses have been used for organizing materials on the nanoscale (Fischlechner and Donath 2007). Finally, and perhaps representing the most fundamental level of materials science and engineering, DNA nanotechnology aimed at building complex nanostructures with multiple functionalities is becoming a reality. Such structures require DNA branching to form 3D, DNA lattices to construct and orient large molecules (Douglas et al. 2012).

Viruses: Structure and Function

Since the discovery of the tobacco mosaic virus by Martinus Beijerinck in 1898, there have been more than 5,000 viruses described in some detail of the millions of virus types which exist. A virus consists of single or double strands of either RNA or DNA surrounded by a protein shell consisting of antigens (lipids) on the surface.

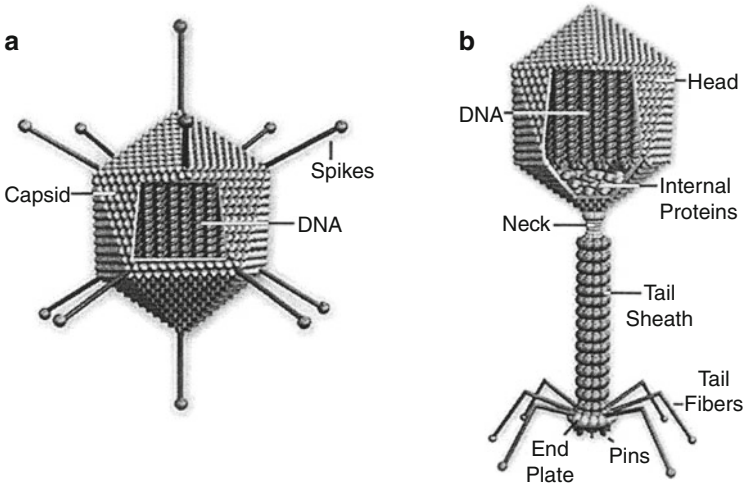


Fig. 4 Virus structures. (**a**) Icosahedral animal virus. (**b**) Bacteriophage (Courtesy of Michael W. Davison, The Florida State University)

While viruses have some characteristics of living organisms, Rybicki (1990) has described them as organisms at the edge of life. Most viruses are about one-hundredth the size of bacteria. Viruses cannot function without a host cell. Outside the cell, the virus is simply an encapsulated RNA or DNA nanoparticle called a *virion* which contains at least one unique protein synthesized by specific genes in its nucleic acid. Viruslike organisms which contain only nucleic acid and have no structural proteins are called *viroids*, while other viruslike particles called *prions* are composed mainly of a protein integrated with a small nucleic acid molecule. There are essentially two shapes of viruses: rods (or filaments) and icosahedral polyhedra. The rod or filament shape is due to the linear arrays of nucleic acid and protein subunits composing the protein shell or capsid which encloses the nucleic acid. Figure 4a illustrates this structure for an animal virus which parasitizes living animal cells. In contrast, Fig. 4b illustrates a bacteriophage which is a virus that parasitizes bacteria; that is it infects and replicates within bacteria. In this sense, they are antibacterial agents and have in fact been used in Russia and parts of Europe as antibiotic alternatives (Keen 2012).

Virus survival requires that they reproduce inside living cells. To achieve this, the virus attaches to the host cell surface at specific receptors using proteins on its surface. The virion then enters the cell by being engulfed by the cell membrane or by fusing into the cell membrane. Inside the cell the virus sheds its protein shell (Fig. 4a) and allows its genetic material to take over the host cell functions to make millions of new virus particles using cellular material. Protein shells for new virus particles are programmed from a copy of the genetic material. Once the new virus shells are formed, the virus particle leaves the cell either by rupturing the cell membrane which destroys the cell or by slowly budding out from the cell membrane surface.

In the case of bacteriophage illustrated in Fig. 4b, they replicate within bacteria following the injection of their genome into the cytoplasm. Like bacteria, plant cell walls are thicker than animal cell walls and more difficult to breach. Virus self-assembly within host cells has implications for the more general concept of organic molecule self-assembly from a fundamental materials perspective since viruses can be regarded as organic nanoparticles. In this sense, viruses can be used as scaffolds for covalently linked surface modifications (Fischlechner and Donath 2007).

Bacteria: Structure and Function

Figure 5 illustrates common bacterium structures and functional features in contrast to those illustrated in Fig. 4 for viruses. It can be noted that DNA strands occupy the central regimes of both viruses and bacteria, although the length scales are fundamentally two-orders-of-magnitude different. Unlike viruses which are transported tactilely or by airborne migration, most bacteria achieve motility (in a fluid environment) by flagella which are characteristically nanomotors powered by ionized hydrogen or protons (H^+) as illustrated schematically in Fig. 5c. In some bacteria, Na^+ ion rather than H^+ creates the electromotive gradient. In the case of *E. coli* shown schematically in Fig. 5a, six flagella are connected to the rotating hook, while in other organisms the hook is connected to a single filament or flagellum (single-pole flagellated bacterium) which rotates with a screw-like motion to propel the organism toward energy sources (nutrients) recognized by chemoreceptors or chemosensors, a process called chemotaxis. The motor direction can also be reversed when the chemoreceptors detect toxic substances or harsh environments. Rotation rates of the flagella motor (Fig. 5c) can vary from ~300 to 1,700 Hz with torques around 4×10^{-18} Nm (Minamina et al. 2008). This propels the organism roughly 35–115 μm/s. The flagellum is made up of ~25 different proteins representing 1 % of the bacterium's total protein and requires ~50 genes for its production, roughly 2 % of the genome of ~2,500 genes.

The features of the flagellum and flagellar motor shown in Fig. 5c have emerged from high-resolution transmission electron microscopy and cryomicroscopy which allows tomographic images or "slices" through the organism to be reconstructed into 3D images of the entire cell (Suzuki et al. 2004). The flagellar motor is self-assembled forming a stator-rotor structure implicit in the protein configurations shown in Fig. 5c (McNab 2003). *E. coli* cells (Fig. 5a, b) can duplicate their DNA strand and split into two new cells (binary fission) every 20 min.

As illustrated generally in Fig. 5a, bacteria often have pili (or cilia). One type of pili is used by the organism to exchange genetic material, a process called *conjugation*. A second type of pili acts as hold fasts which anchor the organism to a host cell which acts as an important mechanism in infection. In addition to the DNA nucleoid shown in Fig. 5a, bacteria also contain ribosomes and plasmids within the cytoplasm. These are protein inclusion bodies containing nucleotide RNA in the ribosomes and additional DNA genes in the plasmids. As illustrated in the SEM view of *E. coli* in Fig. 5b, there are no pili or flagella as shown in Fig. 5a. This is

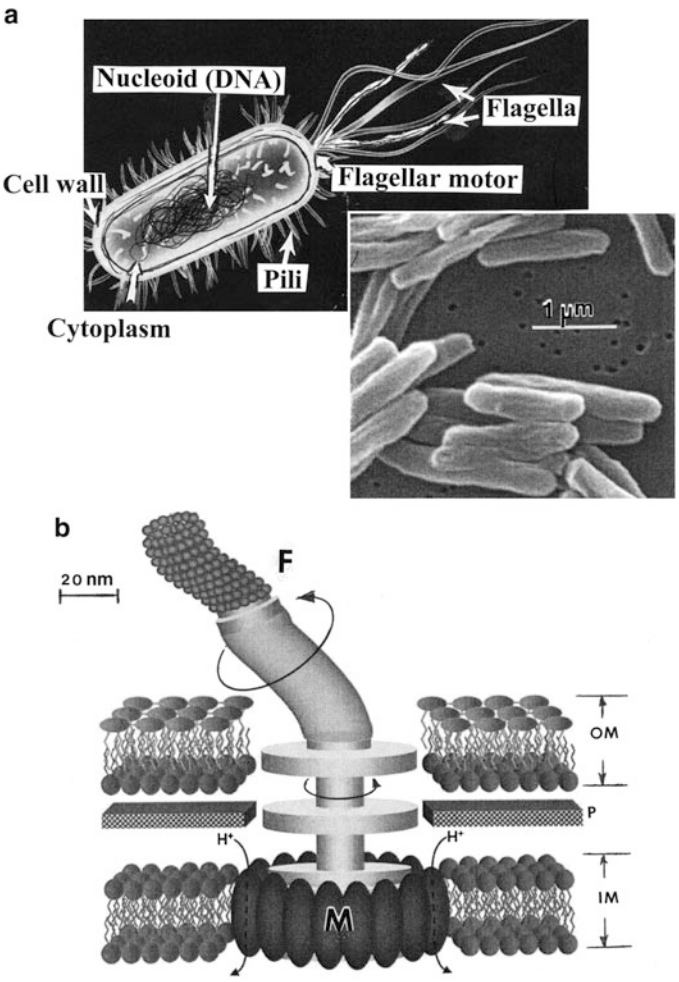


Fig. 5 *Escherichia coli* bacterium. (**a**) Cartoon showing basic bacterial structure. (**b**) SEM of *E. coli* without pili or flagellum which are unpreserved. (**c**) Bacterial flagellum motor structure. *F* represents the flagellar filament. *OM* indicates the outer membrane, *IN* inner (cytoplasmic) membrane. Periplasm (*P*) mesh of very strong fibrous peptidoglycan (murein). Motor proteins (*M*) conduct electric current carried by protons (H^+) from the periplasm into the cell cytoplasm. The motor ring encased in the protein ring (*M*) converts the electric charge into rotary mechanical motion and rotating the flagellum (Adapted from botrejectsinc@cronedon.com)

because without critical point processing (drying) prior to SEM observation, some degree of hydration in these components promotes surface tension effects which collapse them into the surface structure and render them absent.

In some bacteria such as the tuberculosis pathogen as shown in Fig. 6a, the cell wall is an almost impenetrable, hard, lipid polymer. Antibiotics such as streptomycin (Fig. 2) can only kill the organism after the cell wall is breached. In addition, the

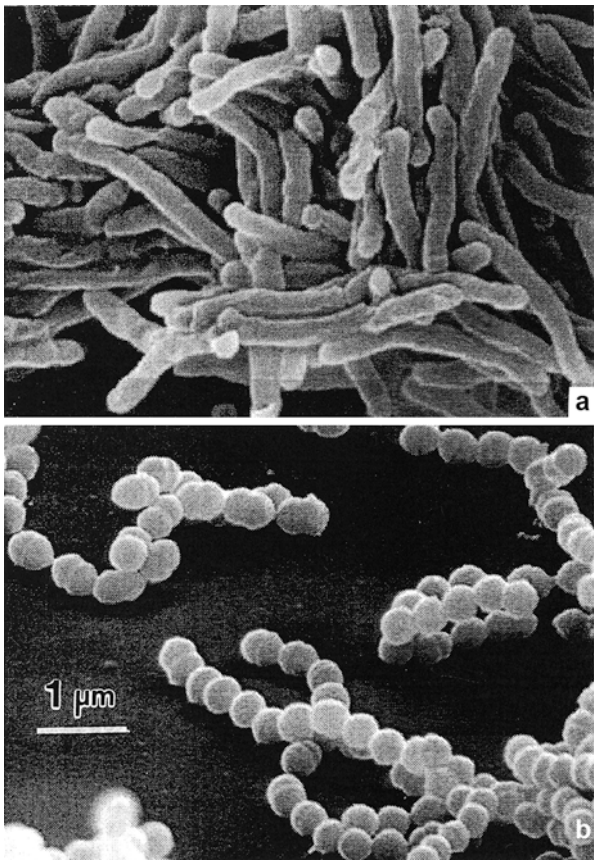


Fig. 6 Rodlike tuberculosis pathogen (*Mycobacterium tuberculosis*) (**a**) compared to streptococcus bacterium (*Streptococcus pyogenes*) a Gram-positive bacterium (**b**) (Google images)

organism mutates against antibiotics by altering the outer cell wall structure. Recently, novel regimens have been developed where an aerosol surfactant sprayed into the lungs of tuberculosis patients effectively breaches the outer cell coating which when followed by a heavy dose of antibiotic effectively kills the pathogens (Stoops et al. 2010).

In contrast to the rodlike (Bacilli) bacteria shown in Figs. 5b and 6a, spherical shapes or connected chains of spheres characteristic of Cocci-type bacteria can occur as illustrated for streptococcus in Fig. 6b. These cells have no flagella and therefore are limited in their motility. In this regard, the tuberculosis pathogen shown in the SEM image in Fig. 6a, like the *E. coli* in Fig. 5b, also shows no pili or flagella as a consequence of hydration-induced surface tension effects which fold these features into the cell surface.

Some bacteria such as *C. cresentus* have a single protein layer on its surface. Recently, the system that secretes this self-assembling protein forming a structure called the "S-layer" has been adapted to instead secrete many proteins that are useful for vaccines and related therapeutic purposes (Ackbarow and Buehler 2011; Villaverde 2010; Vasquez and Villaverde 2010). These can include HIV (virus)

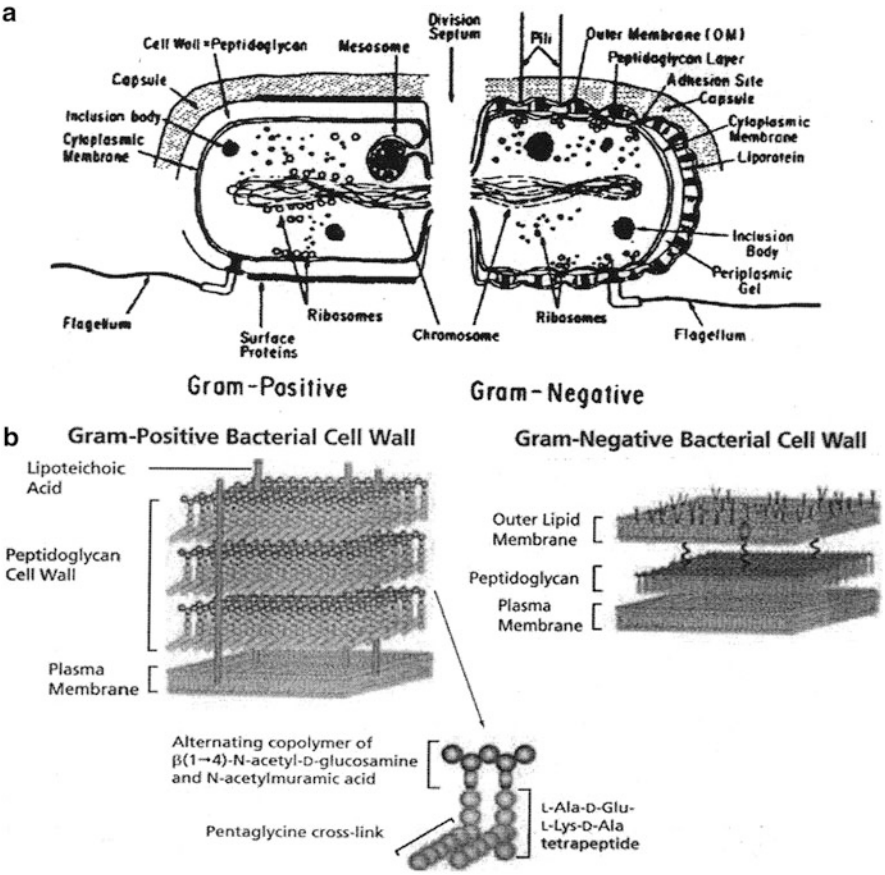


Fig. 7 Bacteria cell wall structure. (**a**) Gram-positive bacteria such as *Bacillus*, *Streptococcus*, and *Staphylococcus*. (**b**) Gram-negative bacteria such as *E. coli*, *Helicobacter pylori*, and *Haemophilus influenzae*. (See Gram-staining protocol) (Google images)

infection blockers and agents to treat Crohn's disease and colitis. This is part of the bacterial cell factory development mentioned earlier (Rodriguez-Carmona and Villaverde 2010).

The cell wall (or envelope) structure for Gram-positive and Gram-negative bacteria is somewhat different, consisting of essentially three layers and a plasma membrane versus two layers and a plasma membrane, respectively, as illustrated in Fig. 7. The cell walls are primarily peptidoglycan (murein) made from polysaccharide chains cross-linked by peptides. As observed in Fig. 7, Gram-positive bacteria cell walls are thick, while the Gram-negative cell wall is much thinner, and obviously this produces differences in susceptibility to antibiotics. In addition to cell protection, shape maintenance (mechanical support), and moving nutrients into the cell cytoplasm, the cell wall can have other novel functions in the context of materials systems and especially materials processing.

While the leaching of copper and its recovery from copper-containing solutions was recorded around 1670 (Taylor and Whelar 1942/1943), the first evidence of microorganisms (bacteria) having a role in copper leaching began around 1960 with *Thiobacillus ferrooxidans*, a motile, single-pole flagellated, Gram-negative, chemoautotrophic, iron-oxidizing bacterium shown in Fig. 8a–c. This organism utilizes CO_2 as an energy source and requires sources of H_2 and phosphates for chemosynthesis and growth through energy derived from the catalytic oxidation of ferrous iron ($Fe^{2\dagger}$), insoluble metal sulfides (Fe, Cu, Zn, etc., sulfides), or elemental sulfur which results from catalytic reactions during leaching as represented in the equation sets shown in Fig. 9. The leaching and catalytic role of *T. ferrooxidans* in pyritic porphyry copper wastes is illustrated in Eqs. 1, 2, and 3 in Fig. 8d, while sulfur production and conversion to sulfuric acid is represented in Eqs. 4 and 5. The catalytic efficiency in these reactions has been demonstrated in the work of Lacey and Lawson (1970) who showed that the reaction rate implicit in Eq. 5 in Fig. 9 is almost a factor of 10^6 times faster for Eq. 6 where the ferrous-to-ferric oxidation can be expressed more generally as Eq. 7, and implicit in the cell envelope function illustrated in Fig. 8c and represented in Eqs. 8, 9, 10, 11, and 12 in Fig. 9.

In leach dump environments, *T. ferrooxidans* become nonviable at ~40 °C, but synergistic or symbiotic catalysis can be achieved by adding extremophiles such as *Sulfolobus acidocaldarius* or other thermophilic bacteria isolated from acidic ($<pH_2$) hot springs (Murr 1980). These acidophilic organisms can extend catalytic leaching to >60 °C. In addition, in pyritic porphyry wastes, contacting chalcopyrite and pyrite (FeS_2) form a galvanic couple. Bacterially catalyzed galvanic conversion of $CuFeS_2$ and passivation of the FeS_2 as illustrated in Fig. 10a, b (where pyrite serves as the cathode and chalcopyrite the anode as illustrated in Eqs. 8 and 9 in Fig. 9) then occur. Figure 10c, d shows experimentally the catalytic effect and temperature symbiosis of *T. ferrooxidans* and *Sulfolobus* in the leaching of $CuFeS_2$/FeS_2 concentrate involving galvanic interaction.

Bacteria that live and thrive in extreme environments such as the *Sulfolobus* or other thermophilic organisms are often referred to as extremophiles as noted earlier. These include acidophiles at $<pH_2$, thermophiles at temperatures between 48 ºC and 120 ºC, halophiles at <0.2 M NaCl, etc. These are often found in hot springs or deep ocean steam vents and the like. Their role in catalyzing metal recovery processes is illustrated in Fig. 10. However, the galvanic corrosion illustrated in Fig. 10 also presents other issues involving microbial corrosion such as pitting corrosion in gas and oil pipelines. Microbial degradation can also occur for polymeric materials as well (Heitz et al. 1996).

Recent research by Bruce Logan at Penn State University has demonstrated that a methanogen strain, *Methanobacterium palustre*, could convert electrical current directly into methane. Methanogens are usually rod bacilli or coccoid as illustrated in Fig. 6. They are anaerobic organisms and can reduce CO_2 in the presence of H_2 to produce methane and water:

$$CO_2 + 4\,H_2 \rightarrow CH_4 + 2\,H_2O.$$

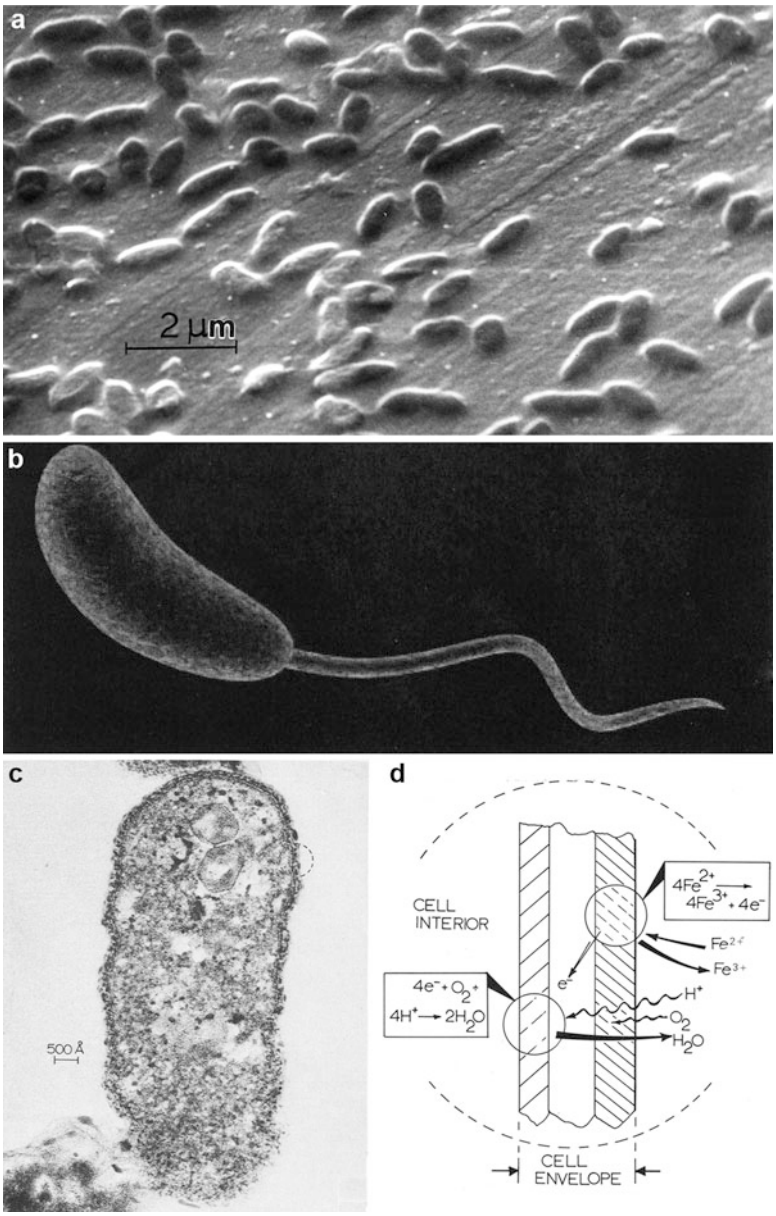


Fig. 8 Bioleaching of copper from chalcopyrite ($CuFeS_2$) with *T. ferrooxidans*. (**a**). SEM view. (**b**) Cartoon illustrating single flagellum. (**c**) TEM image of stained section of *T. ferrooxidans* and schematic representation of the cell envelope region shown circled indicating the iron oxidation mechanism (Adopted from Murr 1980)

Equations	
$FeS_2 + 3.5\ O_2 + H_2O \xrightarrow{\text{bacteria}} FeSO_4 + H_2SO_4$	(1)
$2\ CuFeS_2 + 8.5\ O_2 + H_2SO_4 \xrightarrow{\text{bacteria}} 2\ CuSO_4 + Fe_2(SO_4)_3 + H_2O$	(2)
$FeS_2 + Fe_2(SO_4)_3 \rightarrow 3\ FeSO_4 + 2S$	(3)
$2\ S + 3\ O_2 + 2H_2O \xrightarrow{\text{bacteria}} 2\ H_2SO_4$	(4)
$2\ FeSO_4 + 0.5\ O_2 + H_2SO_4 \rightarrow Fe_2(SO_4)_3 + H_2O$	(5)
$2\ FeSO_4 + 0.5\ O_2 + H_2SO_4 \xrightarrow{\text{bacteria}} Fe_2(SO_4)_3 + H_2O$	(6)
$4\ Fe^{2+} + O_2 + 4H^+ \xrightarrow{\text{bacteria}} 4\ Fe^{3+} + 2\ H_2O$	(7)
$O_2 + 4e^- + 4\ H^+ \rightarrow 2H_2O$	(8)
$CuFeS_2 \rightarrow Cu^{2+} + Fe^{2+} + 2S^\circ + 4e^-$	(9)
$CuFeS_2 + O_2 + 4H^+ \rightarrow Cu^{2+} + Fe^{2+} + 2S^\circ + 2H_2O$	(10)
$M \rightarrow M^{n+} + ne^-$	(11)
$MS \rightarrow M^{n+} + S^\circ + ne^-$	(12)

Fig. 9 Equation set for bacterial leaching of chalcopyrite ($CuFeS_2$) (From Murr 1980)

This creates an electrochemical gradient across the cell wall, and this feature can be reversed. Many methanogens have an S-layer cell wall structure as described above.

Magnetospirillum magneticum represents so-called magnetic bacteria which can often be found in shallow pools of water containing iron. The bacteria draw the iron from the water and create magnetite (Fe_3O_4) nanocrystals within the cytoplasm. They use these nanomagnets to become oriented in the Earth's magnetic field which aids in accessing necessary oxygen sources. *Geobacter sulfurreducens* has been observed to create electrically conductive filaments or bacterial nanowires (Gorby et al. 2006). Consequently the prospect for interaction or related utilization of these organisms elicits innovations in biodevices, including biocomputers.

In addition to materials synthesis and related possibilities posed by manipulations of virus and bacterial cells (including the cell wall), other biological systems are useful in biosorption and bioremediation or bio-recovery. These include various

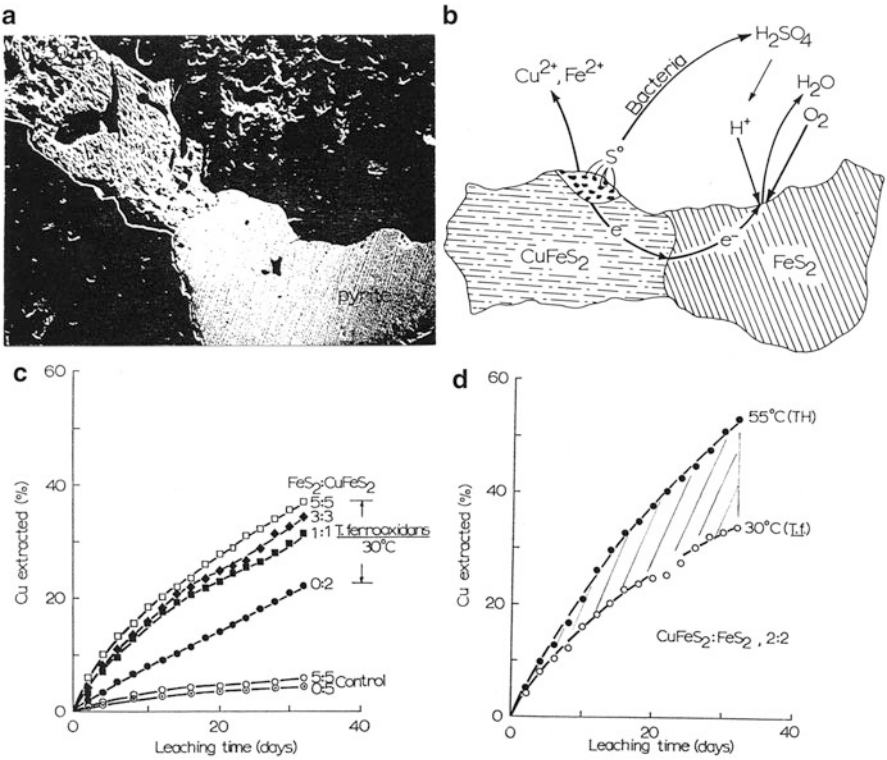


Fig. 10 Contacting chalcopyrite and pyrite showing galvanic corrosion of the anodic $CuFeS_2$ in acid bacterial leaching. (**a**) SEM image. (**b**) Schematic representation of the galvanic reaction in (**a**). (**c**) and (**d**) show copper extraction from $CuFeS_2/FeS_2$ mixtures in ratios shown. *TH* represents thermophilic organisms; *T. f.* represents *Thiobacillus ferrooxidans* (pH 2.3, stirring rate of 100 rpm) (From Murr 1980)

plant materials such as tree bark or the uptake of metal ions in plant roots and stems in particular (Arthur et al. 2005; Murr 2006). The reader is referred to these references for more details.

References

Ackbarow T, Buehler MJ (2011) Strength and robustness of protein materials. In: Nalwa HS (ed) Encyclopedia of nanoscience and nanotechnology, vol 23. American Scientific Publishers, Valencia, pp 349–387

Arthur EL et al (2005) Phytoremediation: an overview. Crit Rev Plant Sci 24(2):109–144

Douglas SM, Bachelet I, Church GM (2012) A logic-gated nanorobot for targeted transport of molecular payloads. Science 335(6070):831–834

Fischlechner MK, Donath E (2007) Viruses as building blocks for materials and devices. Angew Chem Int Ed 46(18):3184–3193

Gorby Y, Yanina S, McLean JS et al (2006) Electrically conductive bacterial nanowires produced by *Shewanella oneidensis* MR-1 and other microorganisms. Proc Natl Acad Sci 103(30):11358–11363

Heitz E, Fleming HC, Sand W (1996) Microbially influenced corrosion of materials. Springer, Berlin

Keen EC (2012) Phage therapy: concept to cure. Front Microb 36

Lacey DT, Lawson F (1970) Kinetics of the liquid phase oxidation of acid ferrous sulfate by the bacterium *Thiobacillus ferrooxidans*. Biotechnol Biocngr 12:29–38

Mcnab RM (2003) How bacteria assemble flagella. Ann Rev Microbiol 57:77–100

Minamina T, Imata K, Namba K (2008) Mechanisms of type III protein export for bacterial flagellar assembly. Mol Biosyst 4(11):1105–1115

Murr LE (1980) Theory and practice of copper sulfide leaching in dumps and in-situ. Miner Sci Eng 12(3):121–192

Murr LE (2006) Biological issues in materials science and engineering: interdisciplinarity and the bio-materials paradigm. JOM 58:23–33

Pauling L, Corey R (1953) A proposed structure for nucleic acids. Proc Natl Acad Sci 39:84–97

Rodriguez-Carmona E, Villaverde A (2010) Nanostructured bacterial materials for innovative medicines. Trends Microbiol 18:423–430

Rybicki EP (1990) The classification of organisms at the edge of life, or problems with virus systematics. South Afr J Sci 86:182–186

Schatz A, Burgie E, Waksman SA (1944) Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. Proc Soc Exptl Biol & Med 55:66–69

Stoops JK, Arora R, Armitage L, Wanger A, Song L, Blackburn MR, Krueger GR, Risin SA (2010) Certain surfactants show promise in the therapy of pulmonary tuberculosis. Vivo 24(5):687–694

Suzuki H, Yonekura K, Namba K (2004) Structure of the rotor of the bacterial flagellar motor revealed by electron cryomicroscopy and single-particle image analysis. J Mol Biol 337(1):105–113

Taylor JH, Whelan PF (1942/1943) The leaching of cupreous pyrites and the precipitation of copper at Rio Tinto, Spain. Trans Inst Min Metall 52:35–71

Vasquez E, Villaverde A (2010) Engineering building block for self-assembling protein nanoparticles. Microb Cell Fact 9:101–105

Villaverde A (2010) Nanotechnology, bionanotechnology and microbial cell factories. Microb Cell Fact 9:53–60

Structures and Properties of Keratin-Based and Related Biological Materials

Contents

Abstract

Keratin, complimentary fibrous proteins to collagen, is represented by two classes or molecular structures, alpha (α) and beta (β), having similar amino acid sequences and biological functions. Prominent examples include wool and silk fibers, avian materials, biological armor (including varieties of scales and shells), and exoskeleton composites.

Introduction

Keratin is a family of structural, biopolymer, fibrous proteins composing epithelial tissue in animals, including human skin, hair, fingernails, animal wool, hoofs and horns, and beaks and feathers of birds, to name the more obvious. Keratin proteins are self-assembled into fibers having four levels of structure with the polypeptide chain forming the backbone. Keratins are composed of amino acids, primarily alanine and glycine, although up to 20 amino acids can compose keratins and other living systems, including cystine and serine as illustrated in Fig. 1.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_28

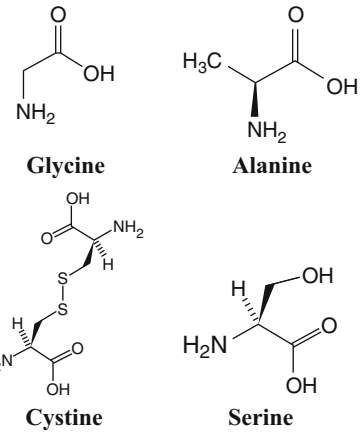


Fig. 1 Principal amino acids composing keratin-based protein fibers. The molecular units for glycine, alanine, and serine are represented somewhat differently from ▶ Fig. 1 of chapter "Structure and Function of Viruses and Bacteria"

Cystine is a dimeric amino acid formed by the oxidation of two cysteine residues (▶ Fig. 1 of chapter "Structure and Function of Viruses and Bacteria") which are covalently linked to form a disulfide bond as illustrated in ▶ Fig. 1 of chapter "Structure and Function of Viruses and Bacteria." In wool fibers which contain 9.8 mol% cystine, the disulfide bridges provide thermally stable cross-linking of the keratin molecules similar to the role of sulfur in the vulcanization of rubber implicit in ▶ Fig. 9c in chapter "Examples of Man-Made Composite Structures".

Keratin is composed of two classes or molecular structures, alpha (α)-keratin and beta (β)-keratin, each having similar amino acid sequences and biological functions. α-Keratin (Fig. 2a) is the major protein in mammalian hair and wool and has a long, alpha-helix molecular structure with its axis parallel to the fiber axis. β-Keratin (Fig. 2b) or fibroin, the major protein in silk, has a pleated sheet structure where identical or opposed molecular strand alignments make up parallel or antiparallel beta-sheets, with the antiparallel beta-sheet being significantly more stable due to the well-aligned H-bonds as illustrated in Fig. 2. These structures, with repeat distances illustrated in Fig. 2a, were originally solved by Linus Pauling as indicated previously (Pauling et al. 1951). In the α-keratin chain in Fig. 2a, the periodic repeat between identically oriented R-groups is 0.7 nm, and this corresponds to a 0.35 nm repeat between amino acids in β-keratin or fibroin (Fig. 2b). Correspondingly, in the α-keratin helix shown in Fig. 2a, the periodic repeat distance separating each turn of the helix is 0.54 nm, and this helical structure then has 0.54 nm/3.6 or 0.15 nm spacing between amino acids.

As to be seen shortly, wool fibers are composed of complex packing of α-keratin helices forming protofibrils, microfibrils, and macrofibrils composing cortical cells in the inner fiber structure which are held together by gamma (γ)-keratin, a globular protein containing fractions of cysteine (Fig. 1) which provides disulfide cross-linking of the α-keratin helical chains. Correspondingly, the β-keratin or fibroin structure composes the primary silk fiber. Table 1 illustrates the amino acid composition of these natural fiber biomaterials.

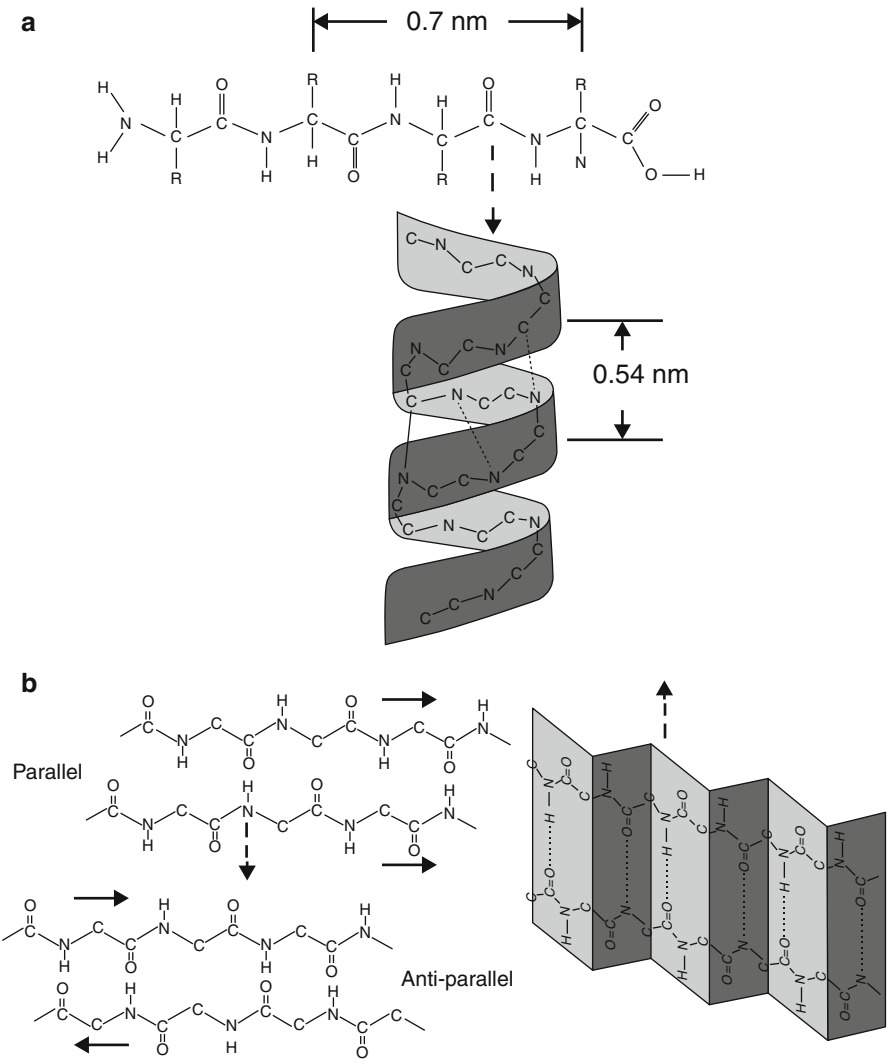


Fig. 2 Alpha-keratin helical structure (**a**) and beta-keratin crystalline sheet structure (**b**). The sheet structure is a pleated sheet structure: parallel or antiparallel. Note in the helical form in (**a**) that only the C–C–N–C–C–N central chain structure is shown. In (**b**), the R-groups (radical components) shown in (**a**) are circled C. Note *dotted arrows* in (**a**) and (**b**) show molecular chain or bond sequence

Wool

As illustrated in Fig. 3, wool fibers consist of an outer layer or cuticle ~l μm thick of overlapping keratin scales somewhat characteristic of butterfly wing scales illustrated in ▶ Fig. 14 in chapter "Electromagnetic Color and Color in Materials".

Table 1 Principal amino acid composition of wool and silk fibers (in mol%)[a]

		Silkworm silk		
Amino acid	Wool	Fibroin	Sericin	Spider silk
Glycine	8.4	44	14.7	37
Alanine	5.5	29	4.3	21
Cystine	9.8	0.2	0.1	0.3
Serine	12	12	37.3	9.2
Other amino acids	Balance	Bal.	Bal[b]	Bal.

[a]Data from Singha et al. (2012)
[b]Includes ~15 % aspartic acid. In contrast, fibroin contains only 1.6 % aspartic acid

The outer cuticle surface is covered by a 3 nm thick hydrophobic epicuticle. Below this thin nanolayer, there is a cysteine-rich exocuticle layer forming roughly two-thirds of the cuticle scale thickness which is followed by the endocuticle and a thin layer of intercellular cement.

The main shaft of the wool fiber shown schematically in Fig. 3b consists of a cortex composed of elongated cortical cells each roughly 2 μm in diameter. Each cell consists of macrofibril bundles which are composed of microfibrils ~7 nm in diameter which in turn consist of a so-called protofibril. A protofibril, measuring ~2 nm in diameter, is formed when three right-handed α-keratin helices (Fig. 2a) are wound together like rope forming the triple helix structure originally described by Pauling et al. (1951). Then, in a 9 + 2 arrangement, the protofibrils combine to form a microfibril. Gamma keratin forms the matrix that holds the microfibrils together and is high in sulfur as a consequence of its cystine content. In fact, the wool fiber surface, which contains some wax or fatty acids, illustrates the disulfide binding in Fig. 3c. Actually, the fatty acids which are on the wool surface are primarily wax, and in wool production for textiles, this wax is removed in hot water (>40 °C) and lanolin extracted from this product. Human hair is similar to wool structure but varies in the ratio of alpha-to-gamma keratin. Human hair is also much larger in diameter than wool fibers and ranges from 100 to 300 μm in contrast to 20–30 μm for wool fibers.

Because of the cuticle scale structure of wool as illustrated in Fig. 3a, individual fibers stand apart when they are combined in forming knitting rope. In a precursor carding process, the wool is combed to straighten and untangle the fibers, aligning them parallel to one another prior to forming yarn (knitting) rope. The irregular wool surface features allow air to be trapped within and between fibers. This layer-trapped, still air provides outstanding insulating properties and is optimized in sheepskin and wool pile which insulate against both heat loss and heat gain. Wool also absorbs and wicks moisture, and individual fibers can absorb up to 34 % of their weight in water vapor. In contrast, synthetic polymeric fibers can only hold up to ~2 % moisture.

A novel fiber composite concept applicable to wool and other fibers involves the infusion of nanosilver or colloidal silver particles (▶ Fig. 10b in chapter "Examples of Materials Science and Engineering in Antiquity"). This is especially viable in the case of wool because of the ability of the fiber scales (Fig. 3a, b) to trap the

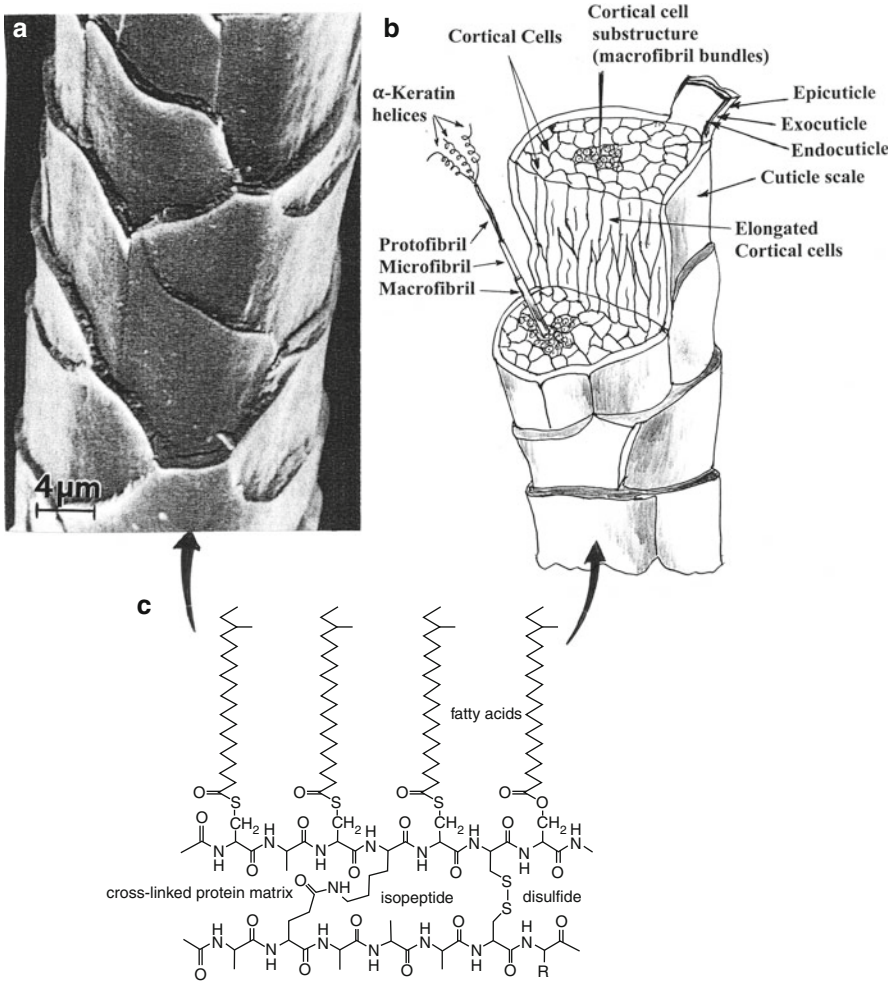


Fig. 3 Wool structure. (**a**) SEM image of wool fiber section. (**b**) Structural hierarchy. (**c**) shows the chemical structure of wool (cuticle) surface. The fatty acids represent wax which is removed in processing the wool and is a source of lanolin in sheep wool in particular

nanosilver particles. These nanosilver-infused fibers are woven into a variety of fabrics to create antimicrobial and antifungal products which include socks, sportswear, hospital bedding, and a host of related applications. Barani et al. (2012) have recently described enhanced diffusion of nanosilver ions into wool fibers and the control of the release rate of the nanosilver particles or ions. Similar infusion processes have been developed for other polymer fibers such as nylon, and nanosilver solution treatment for surface finishing on cotton woven fabric has also been developed. Johnston and Nisson (2012) have recently described nanogold and nanosilver composites with lignin-containing cellulose fibers in the development of

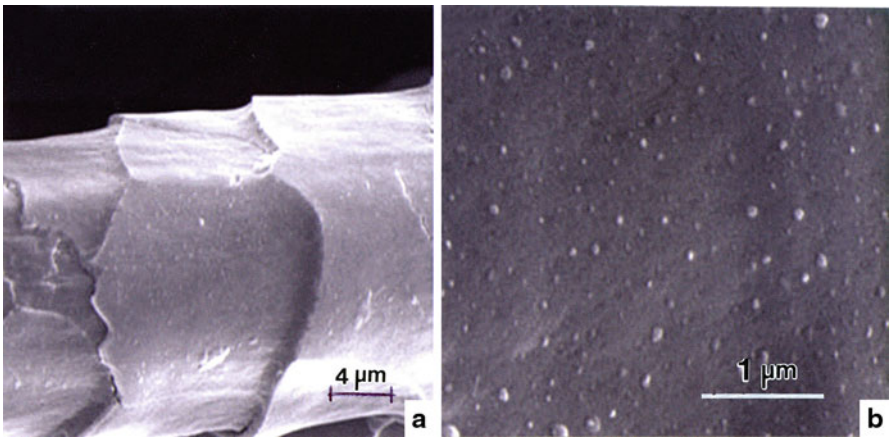


Fig. 4 SEM images of wool fiber (**a**) having 100 ppm Ag particles with sulfur nanosilver colloid. (**b**) shows a magnified view of (**a**) (Adapted from Perumalra (2012))

new functional paper products, including antimicrobial packaging concepts. Similarly, Perumalra (2012) has described the attachment of nanoparticle silver to dyed wool fibers as illustrated in Fig. 4.

It is observed that surface electrical resistivity of wool and other fibers can be reduced by surface films and conducting particle adhesion as illustrated for silver particles in Fig. 4. The electrical conductivity can also be improved significantly by treatment of fibers such as wool by in situ surface deposition of conducting polymers such as polypyrrole which exhibits high conductivity and environmental stability (Varesano and Tonin 2008). Similar polypyrrole coatings have been applied in the preparation of conducting cellulose and silk fibers (Hosseini and Pairovi 2005). These fibers are usually stronger and more stable than conducting polymer fibers, and treated – conducting natural fibers can be selectively woven into fabric patterns to act as conducting interconnects for flexible electronic platforms integrated into textile products. In this regard, it is possible to integrate fine copper- or silver-conducting wires or carbon fibers into automated textile looms for high-efficiency interconnects to electronic components. Novel arrays of optical fibers and LEDs (light-emitting diodes) have also been integrated into fabrics and fiber products.

Silk

Silk or silk fibers are often considered fibroin protein block copolymer structures because the primary structure of the fibers is composed of repetitive block of glycine and alanine amino acids. This primary structure forms a secondary structure of complex chains of β-keratin, antiparallel pleated crystal blocks (Fig. 2b) connected by a complex array of helical (amorphous) keratin linkages. Fibroin is an insoluble protein composed of 16 amino acids, of which alanine, glycine, and

serine constitute 80 % of the polymer (Table 1). Unlike wool, there is no significant cysteine and therefore no disulfide cross-linking, although as illustrated in Fig. 2b, the β-keratin pleated sheet structures involve hydrogen bonds. Fibroin is produced in silkworm silk forming cocoons of the *Bombyx mori* moth larvae which live exclusively on the leaves of the white mulberry tree. A cocoon is made of a single, continuous silk thread which varies in length from roughly 300 to 900 m. Silkworm silk strands consist of two fibroin filaments encased in a protein called sericin as illustrated in Fig. 5a. The fibroin polymer system composing silkworm silk (Fig. 5) is about 65–75 % crystalline. The crystals are blocks of β-keratin antiparallel pleated sheets (Fig. 2b). These can measure roughly 3×5 nm and are composed of layered, pleated sheet blocks within nanofibrils as shown in the hierarchical structure schematic in Fig. 5b. The unit cell for the crystallites has lattice parameters $a = 0.94$ nm, $b = 0.95$ nm, and $c = 0.7$ nm (along the fiber axis). The β-pleated sheet blocks (crystals) are thought to be oriented at some angle, θ, relative to the fibril (nanofibril) axis, and like reinforcing fibers in classical composites (Part. 7), the fiber strength is improved as θ approaches zero. This is a characteristic of spider silk which generally has improved strength in relation to silkworm silk. The random helix structure connecting these crystal blocks can be considered a springlike analogue, often referred to as a block and spring or block and "string" structure (Termonia 1994). This structure serves as a mechanical model as well as a model for thermal transport in the axial direction of single silkworm silk fibers (Liu et al. 2012). The random helix threads or molecular chains forming the amorphous silk matrix in which the crystallite blocks are embedded are also interlinked with hydrogen bonds, creating the so-called β-turn structure implicit in Fig. 5b (Simmons et al. 1996). In spider silks, this can form semicrystalline regimes which contribute to the structural differentiation of silkworm silk and spider silk (Wu et al. 2009).

The mulberry silkworm was first domesticated in China nearly 5,000 years ago. Sericulture and the processing of silk fibers produced large amounts of silk during the reign of Han Wudi between 140 and 87 B.C. and were carried over the "Silk Road" from Xi'an, Shaanxi Province, in China to the eastern Mediterranean and other major cities of the West. During the spinning of the cocoon, the larva moves its head in a figure-of-eight pattern to create either a mostly spherical or oval structure. After the moth leaves the cocoon, it quickly mates since their life span is only a few days. Females lay their eggs (up to 500 flat eggs) to complete the silkworm life cycle. Roughly 3 mm silk larva hatches from these eggs in the spring and lives in the fresh mulberry leaves where they spin their cocoon.

Because of the irregular, nearly triangular cross section which is characteristic of silkworm silk fibers as shown in Fig. 5a, light strikes the fibers in many different directions giving silk a natural shine along with a soft texture. Silk does not have surface scales like wool. Silk fibers can elongate up to ~25 % and the tensile strength can vary from ~0.4 to 0.9 GPa. It was recognized for its strength in woven fabrics and was in fact used as body armor or bulletproof clothing up to the First World War. Many events have documented the ability of silk fabrics to stop bullets; prominent among them are folded silk handkerchiefs in the breast pockets of gamblers in the American West which prevented their death during close-range gun battles over disputed gambling winnings.

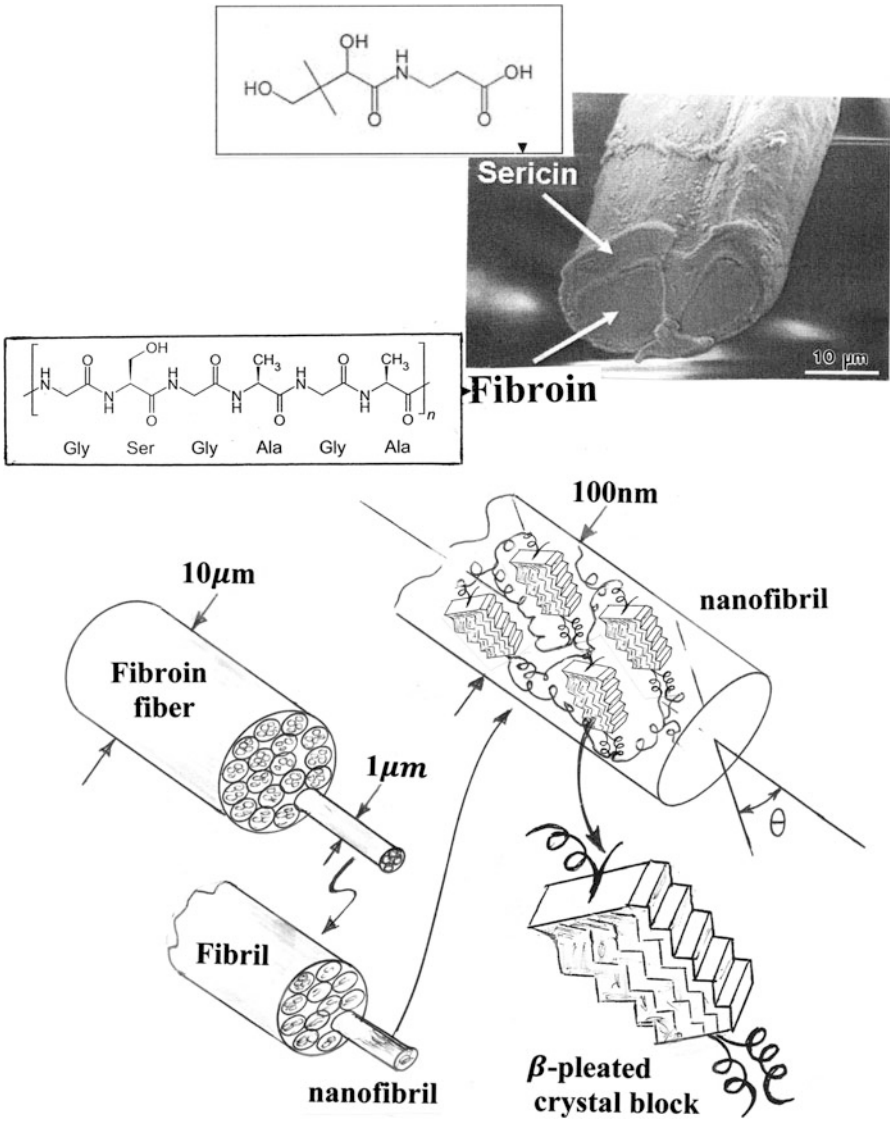


Fig. 5 Silkworm silk structure. (**a**) SEM images showing two fibroin strands surrounded by a layer of sericin. *Insets* show primary fibroin amino acid sequence and sericin structure. (**b**) Fibroin structural hierarchy showing silk fibrils ~1 μm in diameter composing the fibroin fiber. The fibrils are composed of bundled nanofibrils roughly 100 nm in diameter containing arrays of assembled β-pleated sheet crystals linked by irregular (amorphous) keratin helices

In addition to *Bombyx mori* and other moth genera, silk is produced by other insects, especially spiders (Fig. 6). However, spider silk refers to a wide range of continuous, single filaments in contrast to the double-stranded and sheathed silkworm silk illustrated in Fig. 5a. These single fibroin fibers are 2–4 μm in diameter,

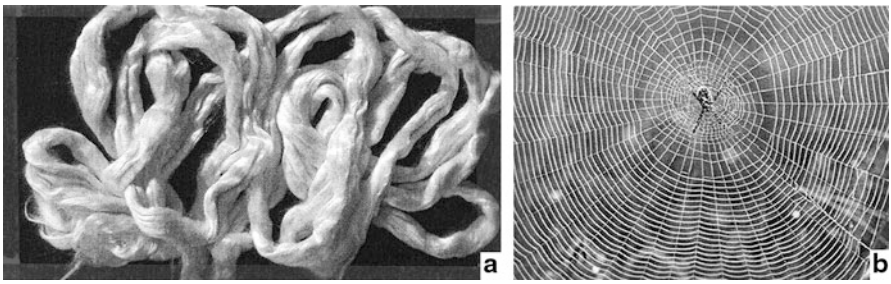


Fig. 6 Comparison of silkworm (mulberry) silk ropes (**a**) and spider silk forming a spider orb web (**b**)

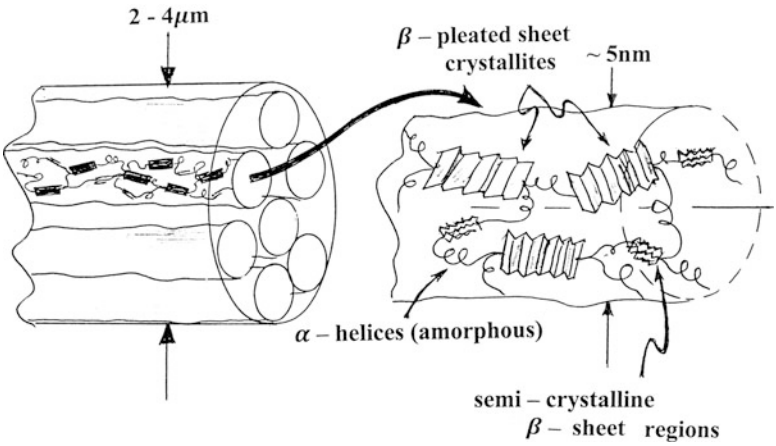


Fig. 7 Structure for spider silk fiber and fibrils. The semicrystalline β-pleated sheets add additional stiffening in contrast to the β-pleated sheet crystallites which compose wool fibrils in Fig. 3b. These are only half the crystallite size, while the spider silk β-pleated sheet crystallites shown are larger than those in wool fibrils

roughly one-tenth the diameter of silkworm silk. Spider dragline silks, like silkworm silk, are composed of highly ordered alanine and glycine polypeptide chains characterizing β-pleated sheet crystallites along with semicrystalline β-sheet regions and α-helical regions rich in glycine in contrast to the β-pleated sheet blocks rich in alanine. These are connected to β-turns or helical strands as illustrated in Fig. 7. In addition, the β-crystallites in spider silk are slightly larger than those in silkworm silk and they tend to be oriented at a smaller angle relative to the fibril axis (Fig. 5b), which provides for a higher elastic modulus and tensile strength in contrast to silkworm silk. Because of the small fraction of spider silk associated with spider webs in contrast to the huge quantities of silkworm silk processed worldwide (Fig. 6), spider silk fabrics are essentially nonexistent. However, 82 people in Madagascar worked 4 years to collect over one million golden orb spiders and

extract silk from them. This culminated in the fabrication of the largest known piece of cloth made of spider silk measuring 3.4 m by 1.2 in 2009 and having a golden tint (see http://www.wired.com/wiredscience/2009/09/spider-silkl).

Spiders can produce up to seven different silk strands depending upon function: in making webs, anchoring the webs, forming nests or cocoons for their offspring, and suspending themselves or for dynamic kiting where they extrude several threads into the air and let themselves be carried away by winds. These very fine silk threads are referred to as *gossamer*. The stronger threads in web making are the dragline silks which have a tensile strength of roughly 1.3 GPa in contrast to 0.5–0.9 GPa for silkworm silk, and 1.8 GPa for nylon fibers, which is half the tensile strength of Kevlar (~3.6 GPa). Silk, including spider silk, has a density of ~1.3 g/cm^3 in comparison to 1.8 g/cm^3 for Kevlar fibers. Additionally, the Young's modulus for dragline (spider) silk varies between 12 and 28 GPa (Keten et al. 2010) depending upon the β-crystal size (Fig. 5).

Among the interesting contrasts between silkworm silk and spider silk is the remarkable difference in thermal conductivity. In single silkworm silk fibers, the thermal conductivity varies from ~0.5 to 6.5 $Wm^{-1}\,K^{-1}$ but when stretched to ~68 %, this value doubles to 13 $Wm^{-1}\,K^{-1}$ (Liu et al. 2012). However, the thermal conductivity in the silk cloth direction (sheet thickness direction) is only 0.042 $Wm^{-1}\,K^{-1}$. Consequently, silk fabrics keep warm air close to the skin during cold weather. In contrast, the thermal conductivity of spider silk fibers is as high as 416 $Wm^{-1}\,K^{-1}$. This exceeds silkworm silk by roughly a factor of 60 and is slightly more than the thermal conductivity of copper (401 $Wm^{-1}\,K^{-1}$).

Avian Materials and Structures

Feathers and beaks of birds illustrate different structural forms of β-keratin pleated sheets which facilitate a number of functions including flight, thermal insulation and water repulsion, camouflage, and novel courtship, including structural color variations. While feathers can range in size and function, for example, very large birds such as the condor and similar large birds such as the emperor penguin (which do not fly and whose small feathers function primarily as insulation), they are hierarchically similar. Figure 8a shows a common pigeon quill or wing feather consisting of a rachis or central shaft which runs the entire length of the feather, with vanes or barbs extending to either side. The right or shorter barb portion in Fig. 8a represents the leading edge of the wing and involves outer vanes. The opposite or left side of the quill in Fig. 8a is constructed of inner vanes. Extending from the barbs is a series of short barbules which create an essential aerodynamic surface by having barbules with hooklets in one direction, whereas in the opposite direction the barbules are ridgelike (Fig. 8b). This opposing hooklet and ridge structures promote attachment of adjacent barblets, a barblet ridge-and-hooklet system. The rachis structure is a complex composite structure consisting of a central closed-cell keratinized foam with a cell size (diameter) of ~10–15 μm which is surrounded by an inner layer of axial fibers of β-keratin ~6 μm in diameter forming

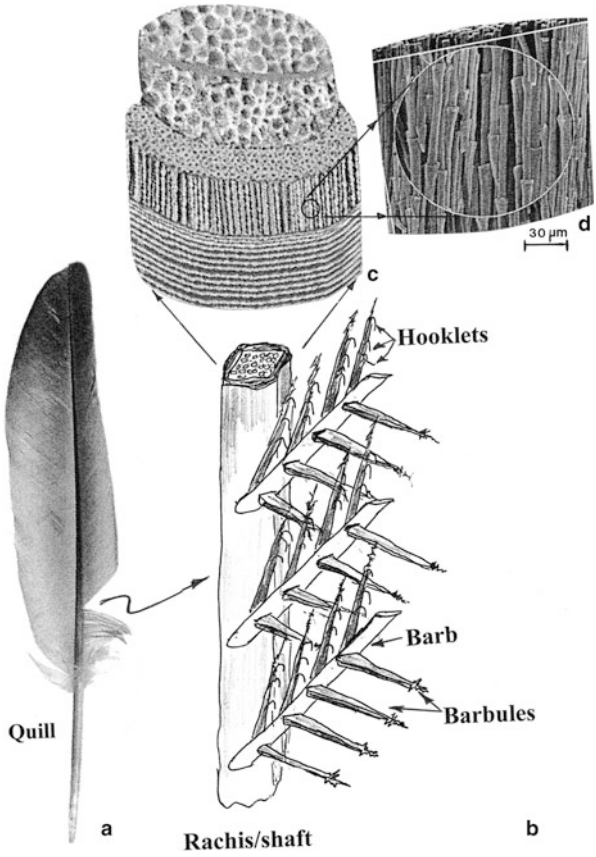


Fig. 8 Bird feather structure. (**a**) Pigeon feather (quill). (**b**) Feather structure details. (**c**) Rachis structure showing closed-cell interior and circumferentially oriented surface fibers. (**d**) SEM image showing aligned fiber structure after removing matrix material by fungal degradation ((**c**) and (**d**) are adapted from Chen et al. (2012))

this inner layer ~250 μm thick, which is surrounded (wrapped) by circumferential fibers (at 90° to the axial fibers) forming a thin surface layer ~15 μm thick. These structural features are illustrated in Fig. 8c, d. The internal cortex, closed-cell foam wall structure is constructed of β-keratin nanofibers ~200 nm in diameter. This multilevel keratinized rachis structure provides light weight and high stiffness (and fracture toughness of 10 KJm^{-2}) somewhat characteristic of bone, where the central cortex foam has a low modulus (~2.5 GPa) while the outer, more dense, harder cortical regime has a higher modulus, ~8 GPa. It can also be noted that the feather rachis structure is fundamentally similar to the hierarchical structure of wood cells (▶ Figs. 8 and ▶ 9 of chapter "Examples of Natural Composites and Composite Structures") and bone and teeth (▶ Figs. 1, ▶ 4, and ▶ 5; see chapter "▶ Examples of Natural Composites and Composite Structures) which embody the harmonious integration of materials and structure.

It is of interest that as insulators, feathers are more efficient than fur, such as wool having the same β-keratin molecular structure. As a consequence, it is apparent that this difference results solely by structure or structural issues.

In considering the insulating properties of feathers, the more remarkable example involves flightless birds: penguins (Fig. 9a). In this context, the emperor penguin, a bird, is the only creature able to survive the Antarctic ice cap in winter, the coldest place on earth. Of the 17 species of penguins, the emperor penguin (*Aptenodytes forsteri*) is the largest, standing 1.1 m and weighing 27–41 kg. Penguin feathers provide excellent insulation in both air and water. They are not arranged in tracts like other birds, but are comprised of an outer "pennaceous" or vane region and a "downy" inner after-feather. The feather shaft has an attached muscle able to pull them into a compressed watertight barrier in water and able to erect them on land. The vanes of the penguin feathers overlap like tiles which create trapped air spaces. There are roughly 15 feathers/cm^2 and ~50 barbs/feather each having ~1,250 barblets emerging from the feather shaft or ramus in a spiral arrangement. The interlocking barbules and hooklets create a relatively uniform, trapped air space. This insulating arrangement of feathers is reinforced by a thick layer of fat just beneath the penguin skin and enables the penguins to stand in huddles for weeks on end at temperatures between −40 °C and −60 °C without eating, and maintaining temperatures in huddles as high as 37 °C. Huddling, as illustrated in Fig. 9b, c, is the key to Antarctic winter survival. Ideally, emperor penguins are packed in huddles so tightly that individual movements are almost imperceptible. However, penguins in huddles move collectively in a highly coordinated manner to ensure mobility without altering the packing density. All penguins make small steps 4–10 cm every 30–60 s essentially traveling as a wave through the entire huddle, which over time leads to large-scale huddle reorganization. This allows individual penguins to take turns moving to the warm huddle interior or cycling from the perimeter to the center of the huddle. In 2012, nearly 600,000 emperor penguins in 46 colonies were reported by satellite imaging along the Antarctic coast.

Unlike other penguins, emperor penguins lay only one egg and immediately after, the female leaves for 3 months during the Antarctic winter, leaving the huddled males (Fig. 9b, c) to incubate the eggs. Females can travel distances from 80 to 110 km. Chicks are born only a few days before the females return to the colony.

Like feathers, birds have a variety of beak structures and sizes (short and thick beaks, long and thin beaks, or long and thick beaks), depending on function (picking insects from tree bark, tearing flesh, drawing nectar from flowers, defense, etc.). Seki et al. (2005) have investigated both toucan and hornbill beak materials and structures. These beaks are long and thin and measure roughly one-third of the total length of the birds. The outside beak shell (integument) consists of thin β-keratin polygonal tile layers forming the epidermis. A dense dermal layer separates the hollow foam composed of cylindrical or elliptical trabeculae enclosed by a thin cortical shell or covering. Figure 10 illustrates these features schematically for a wreathed hornbill bill beak. These beak structures were found to have a large bending strength where the internal cellular core serves to increase the buckling resistance of the beak (Seki et al. 2005). The epidermal keratin tile layers are additionally composed of alternating fiber directions in each successive layer

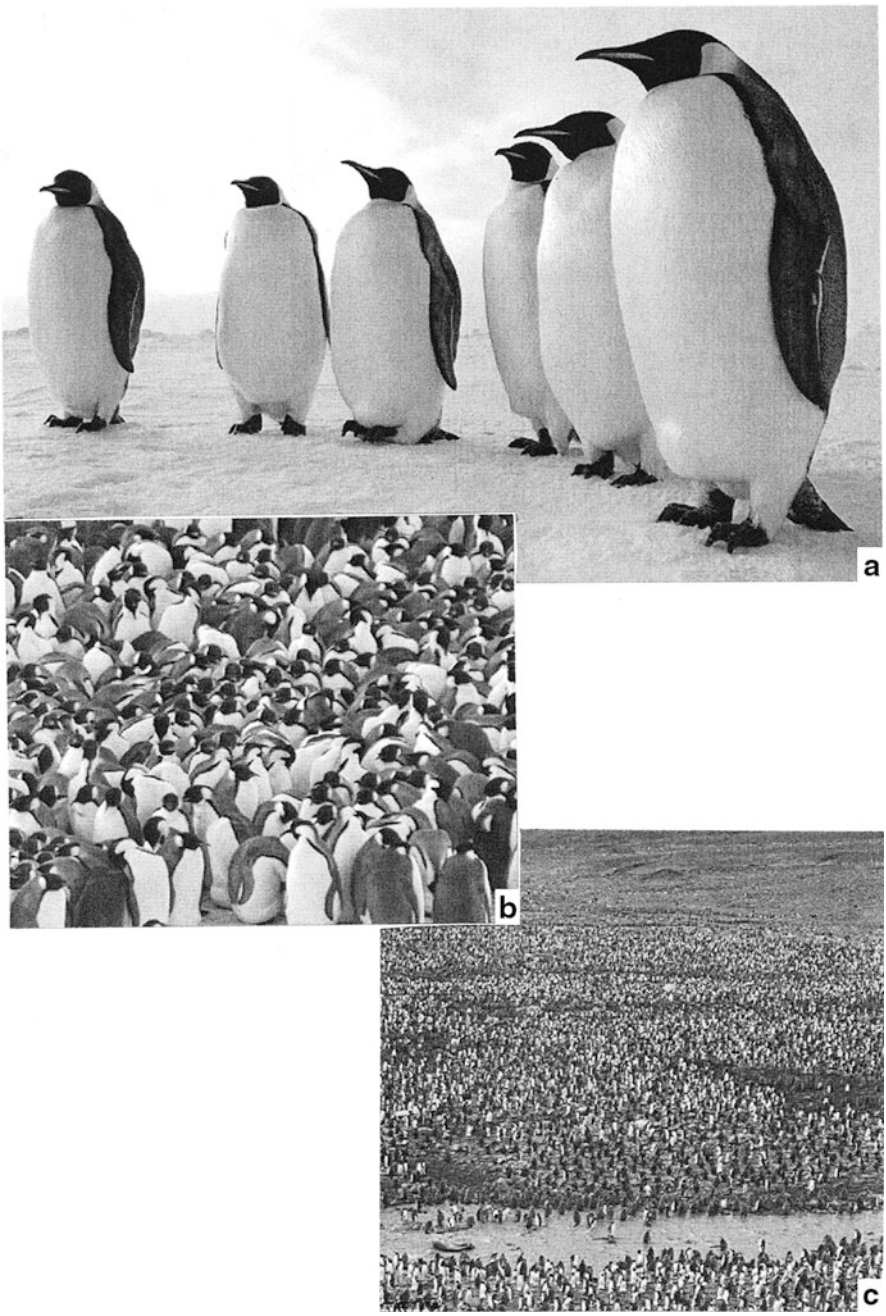


Fig. 9 Antarctic emperor penguins (**a**). (**b**) and (**c**) show penguin huddles which can contain many thousands of birds in a warming configuration (Google images)

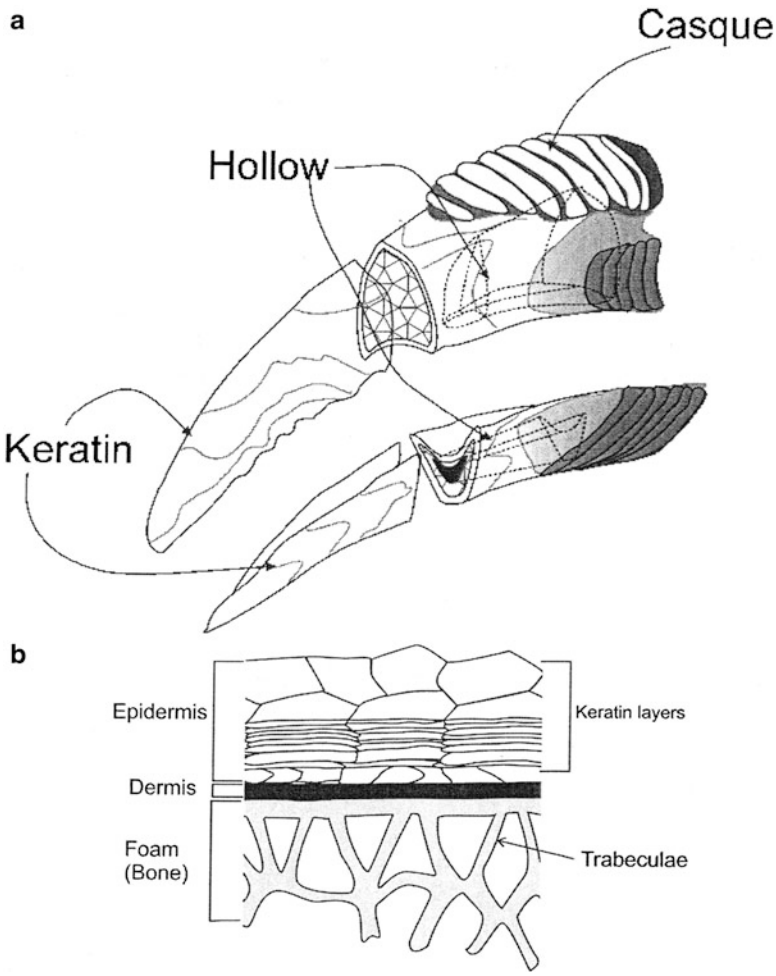


Fig. 10 Schematic views of a wreathed hornbill beak and corresponding beak layers (From Seki (2009))

similar to composite lamina (chapter "▶ Chemical Forces: Molecules and Nanoparticles"). This outer integument structure is relatively ductile in contrast to the semi-brittle bony foam (Fig. 10). The stiffness of the beak keratin is mechanically isotropic in both the transverse and longitudinal directions (Chen et al. 2012; Seki et al. 2005).

While many birds incorporate *melanin* and other pigments or dyes in their feathers to produce color, there are no blue pigments in birds, and blue colors are created by structural elements or nanostructures as described in chapter "▶ Electromagnetic Color and Color in Materials" (Fig. 14c–e). Saranathan et al. (2012) have illustrated that non-iridescent structural colors of feathers are produced by 3D amorphous, spongy β-keratin and air cavity nanostructures found

in the medullary cells of feather barbs. In some bird species such as the six-wired bird of paradise (or Lawes' parotia), the barbules also produce an array of structural colors depending upon the optical angles at which light strikes them and is correspondingly reflected. This bird's barbules are only 5 μm thick and shaped like a boomerang. A thin keratin layer surrounds roughly 25 layers of melanin. As the bird of paradise dances, he displays various colors depending on his position relative to the observing female similar to the movement of a CD relative to a fixed light source and an observer. Such avian amorphous photonic nanostructures have the prospects for biomimetic inspiration for a variety of photonic technologies. These will be explored later in this chapter.

Keratin Armor Examples

One of the simplest protective structures is exhibited by porcupines and hedgehogs which have stiff and sharp α-keratin quills covering their bodies. North American porcupines have roughly 3×10^4 quills on their back. Unlike the quills of birds, these tapered quills have outer scales composed of oriented α-keratin fibrils similar to hair and wool fibers (Fig. 3) which encase a cortical, closed-cell foam structure that maximizes the flexure strength-to-weight ratio. Muscles at the base of the quills allow them to stand up. Hooks at the quill end also allow them to become lodged in a predator, and the flaring of the scales also resists pulling them out.

Scales and shell-related structures are much more prevalent as armor that provides varying degrees of protection against predation and conservation of body fluid. In Squamata (snakes and lizards), a sequence of layers composed of α- and β-keratin forms periodically after which the old epidermal generation is shed. In Crocodilia (crocodiles, alligators, caimans), there is only one layer of β-keratin cornified cells but the layer composition can vary in different parts of the scale. The hinge regions linking the scales contain both α- and β-keratin. In Chelonia (turtles and tortoises), the epidermis of the shell is composed of only β-keratins of turtle shell, whereas the neck and skin is composed of exclusively α-keratin. The β-keratins of turtle shell contain glycine-, proline-, and tyrosine-rich proteins (amino acid residues) which are similar to those in Crocodilia. Table 2 compares the β-keratin amino acid composition for a number of these and related biological regimes. In birds and reptiles, the β-keratins are organized in beta-pleated sheets or strand conformation and aggregated into a densely packed lattice of microfibrils 3–4 nm in diameter.

The upper protective shell of the turtle, the dorsal *carapace*, is a multilayer, sandwich structure. The lower turtle shell or ventral shield is called the *plastron*. The carapace strength and rigidity results from an inner bony casing of fused, connected plates covered by β-keratin scutes or laminae. There are about 38 scutes on the carapace and 12–14 on the plastron. In some turtles, especially captive turtles, the scutes are exaggerated by a pyramidal deformity illustrated in Fig. 11a. The carapace (top shell) structure – also called osteoderms – has a harder, cortical bone on the top and bottom while a cancellous, closed-cell bone structure

Table 2 Amino acid content comparison for human hair, bird feathers, Chelonia, pangolin scales, Crocodilia, and snakes (in wt. %)[a]

Amino acid	Formula	Human hair (keratin)	Bird feather (keratin)	Chelonia (keratin)	Pangolin scale (keratin)	Crocodilia (keratin)	Snakes (keratin)
Glycine	NH2CH2COOH[b]	5.2–6.5	7.2	28	10.6–13.7	30	19
Alanine	CH3CH(NH2)COOH[b]	3.4–4.4	5.4	–	2.3–3.0	–	–
Valine	HO2CCH(NH2)CH(CH32)	5.0–5.9	8.8	–	3.2–4.0	–	–
Leucine	HO2CCH(NH2)CH2CH (CH3)2	7.6–8.1	8.0	–	7.8–7.9	–	–
Isoleucine	HO2CCH(NHY2)CH (CH3)CH2CH3	3.1–4.5	6.0	–	2.0–2.3	–	–
Serine	CH20HCH(NH2)COOH[b]	7.2–9.5	14.0	5.8	5.0–5.5	9.5	12.2
Threonine	HO2CCH(NH2)CH(O4) CH3	6.6–6.7	4.8	–	3.2–3.4	–	–
Tyrosine	C9H11NO3	4.0–6.4	2.2	11	17.2–20.0	10	2
Phenylalanine	C6HSCH2CH(NH2)COO4	3.4–4.0	5.3	–	3.0–3.2	–	–
Cystine	C6H12N2O4S2[b]	11.4–14.1	8.2	–	–	–	–
Cysteine	HOCCH(NH2)CH2SH	–	–	6.5	–	6	5.7
Arginine	C6H14N4O2	9.2–10	7.5	–	8.0–8.6	–	–
Aspartyl[c]	OCCH2CH(NH2)CO	6.4–7.3	7.5	–	7.1–7.9	–	–
Glutamic acid	C5H9NO4	13.1–16	9.7	–	10.4–10.9	–	–
Proline	C5H9NO2	5.8–8.1	10.0	8.8	8.6–9.7	8.9	12.2
Others	–	Bal.	Bal.	Bal.	Bal.	Bal.	Bal.

[a]Data from Tong et al. (1995) and Toni et al. (2007)
[b]See also ▶ Fig. 1 of chapter "Structure and Function of Viruses and Bacteria" and Fig. 1
[c]A bivalent radical derived from aspartic acid

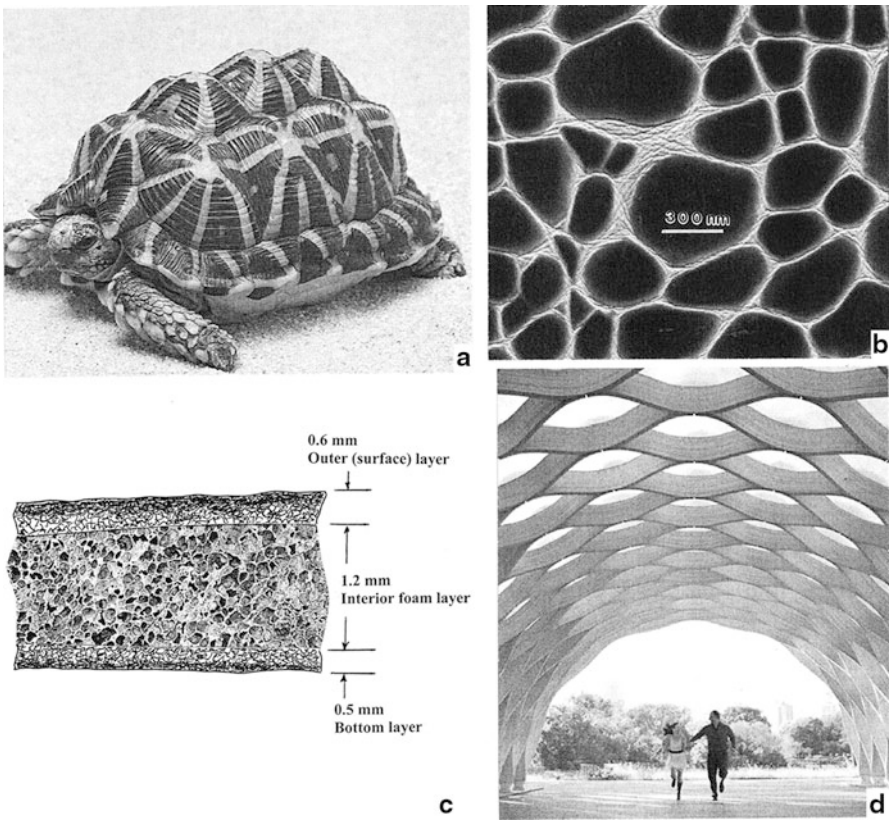


Fig. 11 Chelonia (turtle) shell structure. (**a**) Turtle with pyramidal scutes. (**b**) Cell structure in cross section. (**c**) Shell (carapace) cross section showing layer structure. (**d**) Cellular biomimicry of turtle shell

(foam) is sandwiched between them as shown schematically in Fig. 11b. A cross section of this cellular sandwiched region is shown magnified in Fig. 11c, while Fig. 11d illustrates a biomimetic structure emulating this architecture. The osteoderms have sutures between them composed of non-mineralized collagen fibers which interlock them together. The shell sandwich illustrated in Fig. 11b emulates bone with a harder and higher modulus material (∼20 GPa) sandwiching a soft foam region with a modulus of ∼1 GPa.

Like turtles, armadillos have a carapace divided into five different regions. The top layer of the armadillo carapace is covered with a dark brown β-keratin layer, beneath of which is a well-organized arrangement of bony tiles compacted together and connected by collagen fibers called Sharpey's fibers. Like those in the turtle carapace, they are non-mineralized collagen. Of the roughly 20 species of armadillo which range in length from 0.15 to 1.5 m, the most common armadillo found in the USA is the nine-bonded *Dasypus novemcinctus* which weighs about 6 kg. The armadillo, when threatened, can curl up into a ball-like configuration to protect its head and underbody.

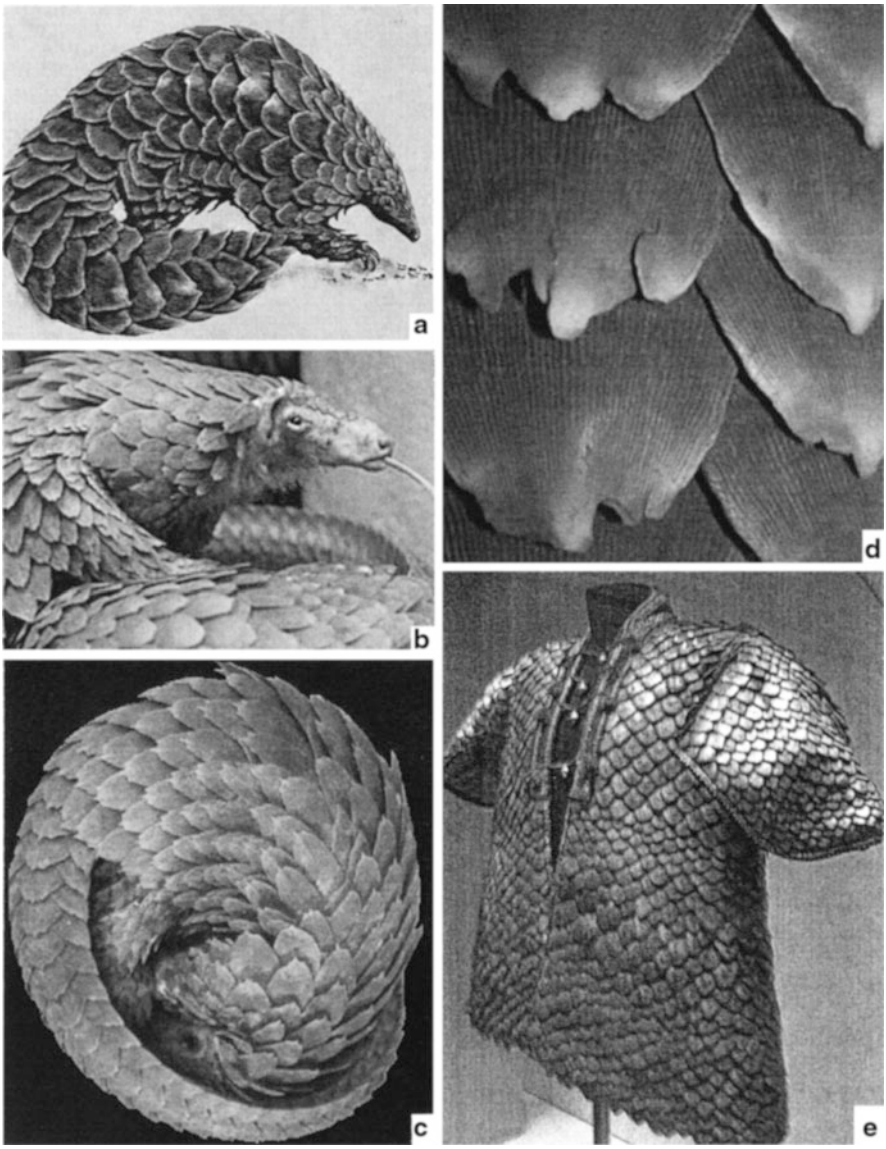


Fig. 12 Armored pangolin (**a**). (**b**) shows enlarged scale structure. (**c**) shows the long sticky tongue for picking up ants or termites. (**d**) illustrates the protective ball position of the pangolin. (**e**) illustrates a pangolin coat presented to King George III which is on display at the Royal Armouries, Leeds, UK

Unlike the armadillo, the Asian and African pangolin is a scaly mammal that feeds on ants and termites. It can be compared to other well-known ant eaters around the world, including the aardvark. The pangolin, as illustrated in Fig. 12, has large β-keratin scales as shown in Fig. 12b which alternately overlap in the regular

scale architecture of many other animals and insects. Its long sticky tongue (Fig. 12c) allows it to capture ants and termites in their nests. They can, like armadillos, curl up into a ball for protection as illustrated in Fig. 12d. A coat of pangolin scales shown in Fig. 12e was created in India in the early nineteenth century and presented to King George III in 1820 by Francis Rawdon, first Marquis of Hastings who was the East India Company's Governor General in Bengal from 1812 to 1822. The scales were decorated in gold leaf. This coat is on display at the Royal Armouries, Leeds, UK.

As illustrated in Table 2, pangolin scales are rich in glycine, tyrosine, arginine, glutamic acids, and proline amino acid residues characteristic of both α- and β-keratin. Tong et al. (1995) have noted in their analysis of pangolin scale behavior that the elasticity and plasticity of α-keratin are greater than those of β-keratin, and the specific elongation of the α-keratin is 100 % while that of the β-keratin is near zero, but the elongation of pangolin scales was observed to be ∼15 %, indicating proportionately more α-keratin than β-keratin. The pangolin scales shown in Fig. 12b have a corrugated surface which arises from oriented α-keratin fibers. These scales present razor-sharp edges when the pangolin is curled up as illustrated in Fig. 12d. α-Keratin also dominates the reinforcement structures for horns and hooves. Figure 13 illustrates a schematic section representing these biological components. The hierarchical structure consists of tubules and laminated structure. Intermediate filaments about 7 nm in diameter consisting of twisted α-keratin helices are embedded in an amorphous matrix, and these filaments extend from one flat lamella to the other. Long tubules extend the length of horns or hooves interspersed between the lamellae. This results in a 3D, laminated composite consisting of crystalline, fibrous keratin embedded in an amorphous keratin matrix. Hoof structures have elliptical tubules in intertubular material with no laminar structure as shown to be characteristic of horns in Fig. 13.

Non-keratin Scale and Insect Exoskeleton Composites

Fish scales provide environmental protection and protection from predators as do the scales of other animals illustrated above. However, fish scales not only possess a novel laminar structure, but also the materials composing this structure are collagen fibers and calcium-deficient hydroxyapatite crystals. These fibers take the form of densely packed lamellae (or lamina) with different orientations as illustrated ideally in Fig. 14. Figure 14a shows an ideal 0°-90°-0° structure element composed of mineralized collagen fibers forming a so-called plywood structure (see ► Fig. 15 in chapter "Examples of Natural Composites and Composite Structures"). Figure 14b shows a TEM image of a cross section of *Pagrus major* (sea bream) scales (Chen et al. 2012), while Fig. 14c shows a similar SEM image showing collagen fibers in the fracture surface of demineralized scales of *Arapaimas gigas*, a freshwater fish in the Amazon basin lakes which can weigh up to 200 kg. Figure 14d shows a more general lamina layer structure also referred to as a Bouligand structure or a twisted plywood structure (Fig. 14a). The collagen fibers composing the layers in Fig. 14a are composed of fibrils made up of collagen twisted strands. The dimensions shown in Fig. 14a change for

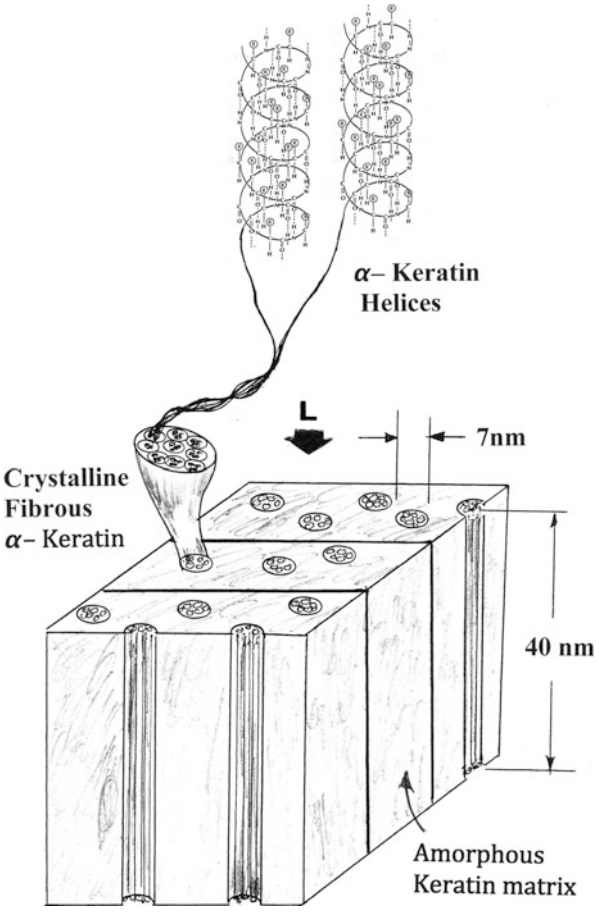


Fig. 13 Schematic views showing general structure hierarchy of horns (specifically bighorn sheep horn). Elliptical tubules of α-keratin embedded in protein-based (keratin) amorphous laminar matrix. *Arrow* marked L illustrates applied compressive load acting on horns or hooves. Hoof structure is similar

various species of fish and the scale dimensions which are large for large fish such as the *Arapaimas gigas*, and becoming very small for very small fish.

Arthropod exoskeleton (cuticle) structure (crustaceans) and insects such as the grasshopper (locust) also have a laminated biological composite similar to fish scales, but somewhat more complex, and composed of chitin fibrils as illustrated schematically in Fig. 15. α-Chitin nanofibers are among the stiffest of natural fibers, having a Young's modulus of ~40 GPa (Chen et al. 2012). Polymer chitin chains illustrated in Fig. 15a form fibrils which form a Bouligand structure in the epicuticle and night layers, while the alternating day layers in the endocuticle are characterized by parallel fibrils as shown in Fig. 15b. The structure also contains pore canals or hollow tubules which join the epidermal cells to the epicuticle and transport waxes to the epicuticle to insure it is waterproof.

Chitin is a glucose derivative, long-chain polymer. As we noted in chapter "► Examples of Natural Composites and Composite Structures" (Figs. 5 and 7), chitin functions as part of the glue to connect aragonite plates in pearls and

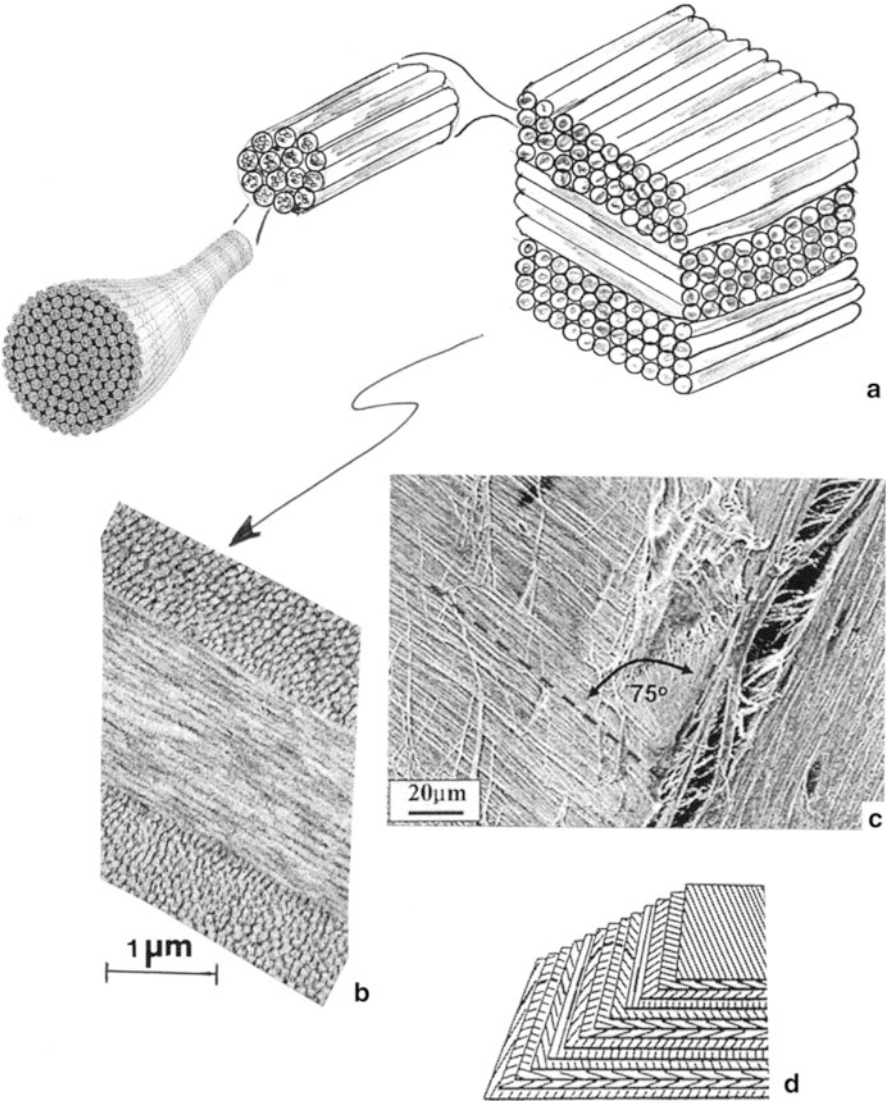


Fig. 14 Composition of fish scales, (**a**) Schematic showing hierarchical structure of scale composed of collagen fibrils forming plywood-like structure fibers (**a**). (**b**) shows a TEM image of a cross section represented in (**a**). (**c**) shows an SEM image of collagen lamellae oriented at 75°. (**d**) shows a more general lamina structure which can represent more complex scales (Adapted from Chen et al. (2012))

Crustacea. In terms of its structure as illustrated in Fig. 15a, chitin may be compared to cellulose, while in the context of function, chitin is comparable to keratin. Figure 16 illustrates this molecular (monomer) structure comparison. The keratin structure shown in Fig. 16c illustrates a primitive molecule which becomes increasingly complex as R-groups are added.

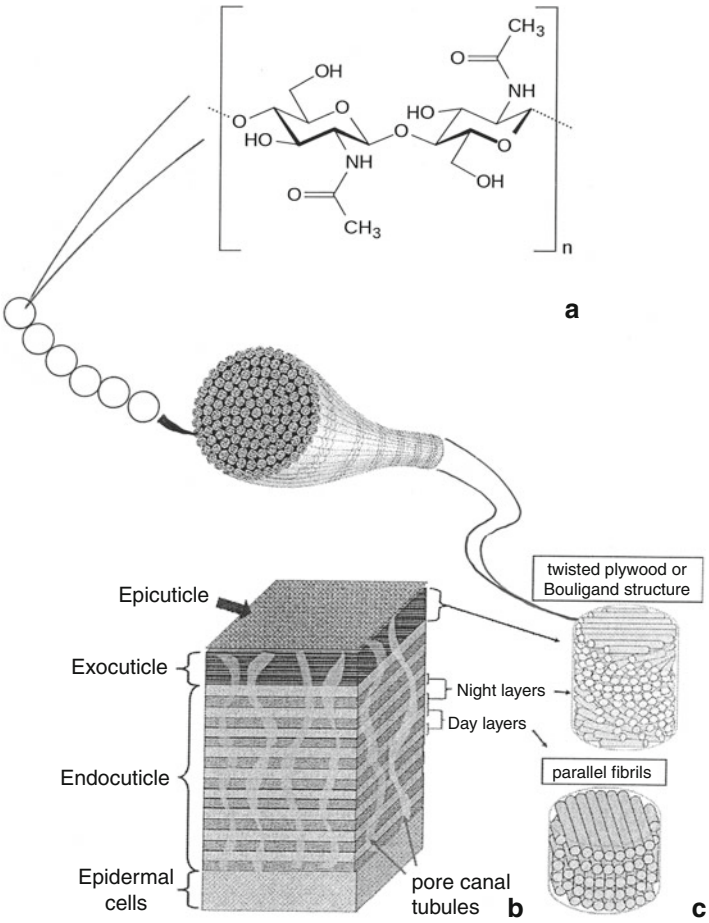


Fig. 15 Schematic representation of insect exoskeleton structural components. (**a**) Chitin molecular monomer forming polymer chain which composes chitin fibrils (**b**) in night and day layers of endocuticle (**c**). Night layer fibrils arrange in twisted plywood structure, while day layers arrange in parallel stacking. Pore canal/tubules are also shown traversing the cell layers ((**c**) is adapted from Chen et al. (2012))

Novel Shell Armor

Although we have discussed composite structures for mollusk shells and pearls in chapter "► Chemical Forces: Molecules," it may be of interest to expand upon this issue as it relates to conch shell (gastropod) structure observed in recent research, especially in the context of armor (Yao et al. 2010). Ballarini and Heuer (2007) have recently illustrated that the queen conch (*Strombus gigas*) not only has crossed lamellar aragonite crystal layers, but the aragonite occurs fundamentally in

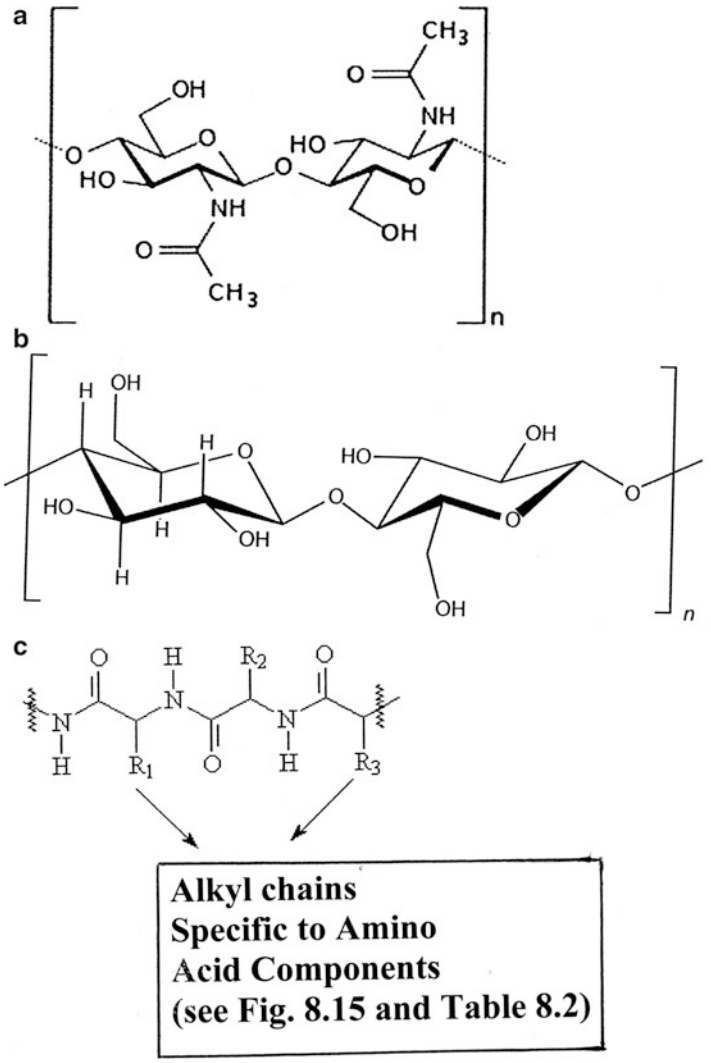


Fig. 16 Comparative molecular structure of (**a**) chitin, (**b**) cellulose, (**c**) keratin

crystalline needles which are coated in a protein sheath and stacked or bundled into beams which in turn are formed into lamellar sheets of aragonite and organic protein matrix. ▶ Figure 17 of chapter "Structure and Function of Viruses and Bacteria" shows a conch shell along with a schematic view of the composite aragonite organic lamellar structure. It can be noted that while the plates alternate the crystallite orientations in each layer of plates, the overlapping plates also crisscross in adjacent plates. This complex, composite architecture has some similarities to the collagen fiber architecture illustrated in Fig. 14a.

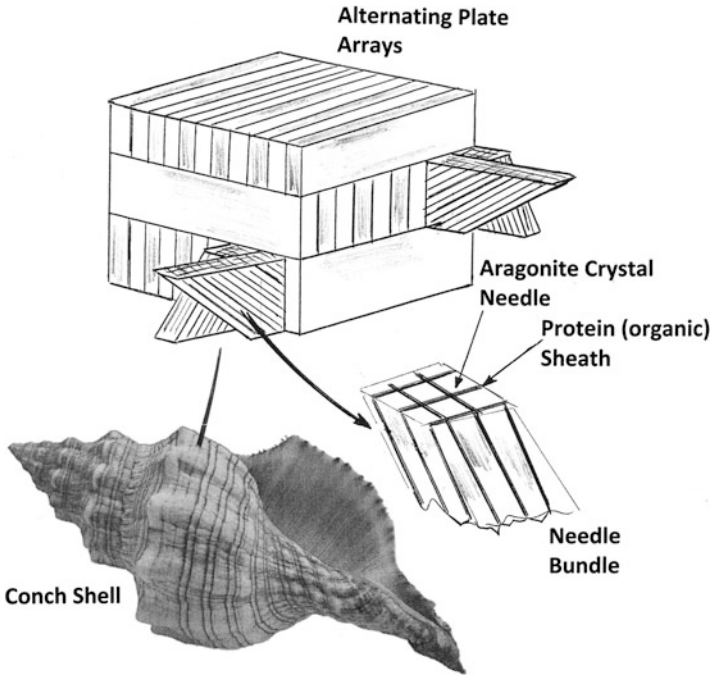


Fig. 17 Conch shell structure. Alternating plate arrays are composed of aragonite crystal needles sheathed in an organic protein (Adapted from Ballarini and Heuer (2007))

The novelty of this complex structure (Fig. 17) is its ability to arrest cracks in addition to its associated strength and toughness. Experiments have suggested that the work of fracture (or toughness) of the conch shell is roughly 1,000 times greater than that of the mineral aragonite by itself.

In a more recent study of a scaly foot snail shell of *C. squamiferum*, a gastropod mollusk discovered at the Kairei Indian hydrothermal vent field, its structure was observed to possess a trilayered architecture comprised of a mineralized iron–sulfide-based outer layer (OL), followed by an organic (conchiolin-based) middle layer (ML) which transitions to a highly calcified inner layer (IL) (Yao et al. 2010). Figure 18 illustrates this curved, multilayer shell armor architecture. The OL in Fig. 18 is ~30 μm thick and composed of Fe_3S_4 composite interface leading to a completely organic region. This region is followed by the IL which is ~250 μm and transitions from organic matter to a calcified (calcite, $CaCO_3$) structure presumably having irregular platelike features. This architecture (Fig. 18) like that for the conch shell in Fig. 17 efficiently dissipates energy of fracture and crack propagation in the layering, especially in the organic matrix of OL and the organic structure of ML.

It might be recalled that the strength of a material is a measure of the force per unit area (or stress, σ) it can resist while the toughness is a measure of the amount of

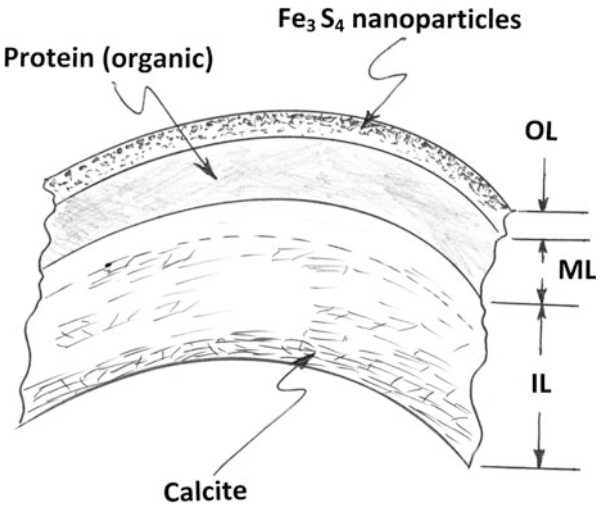


Fig. 18 Schematic view of curved shell section of *C. squamiferum*. The outer shell layer (*OL*) consists of Fe_3S_4 nanoparticles at the outside edge embedded in protein which forms a middle layer (*ML*). The inner shell layer (*IL*) is graded from the protein ML to calcite and in some cases aragonite at the inner edge. The shell curvature also increases exoskeletal stiffness and reduces tensile loads in the inner shell (Adapted from Yao et al. (2010))

work or energy required to break the material or to fracture it catastrophically. Correspondingly, for complex shell composites illustrated in Figs. 17 and 18, toughness is maximized. This is because alternating sheets of lamellae are at right angles or different angles, and cracks therefore can only propagate through the material by following a very irregular path which requires more energy. From first principles, it might be recalled that the Griffith condition stipulates that fracture stress

$$\sigma_F = \left(\frac{2\gamma \mathrm{E}}{\pi c}\right)^{1/2} \tag{1}$$

where γ is the fracture surface free energy, E is the stiffness or Young's modulus, and c is the crack half length and is a characteristic of the crack size. In an engineering sense, toughness of a material generally scales with the area under the stress–strain curve. This represents both the elastic and plastic energies expended. It might be recalled from Part V that we discussed the elastic energy in terms of $\sigma_y\varepsilon_y/2$. Consequently, the total energy is simply related to $\sigma\varepsilon$, where ε can be defined as the true fracture strain: $\varepsilon = \varepsilon_F$. Similarly, the stress at fracture can be defined as $\sigma = \sigma_F$. Toughness, J_c, is then defined as $J_c = \sigma_F\varepsilon_F$. However, the measure of the magnitude of the stress field in the crack tip region is called the stress intensity factor, denoted K or K_{lc}, where

$$K_{lc} \propto (E\sigma_F\varepsilon_F)^{1/2} \tag{2}$$

or

$$K_{lc} \propto (E\, J_c)^{1/2} \tag{3}$$

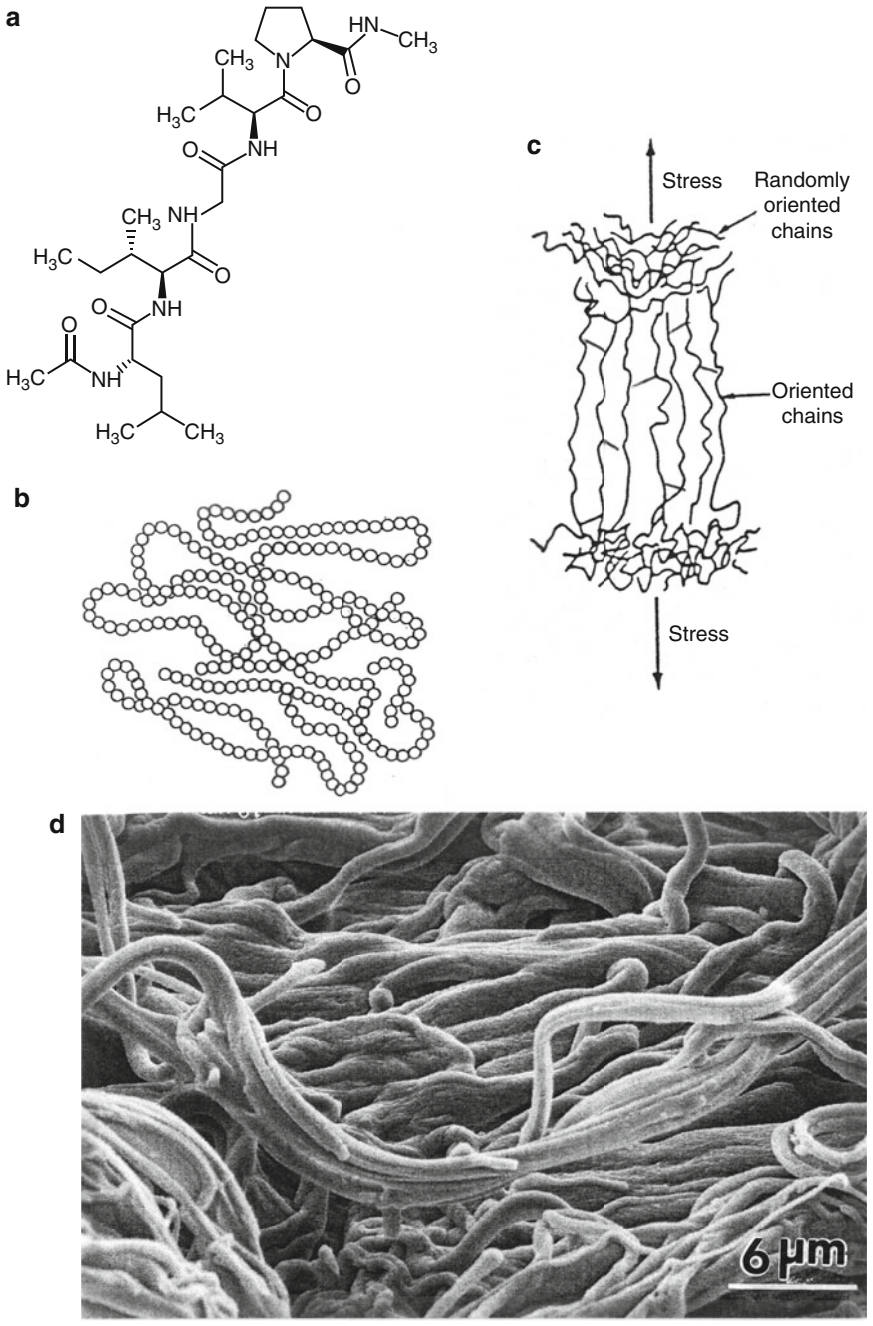


Fig. 19 Elastin. (**a**) Molecular structure unit. (**b**) Schematic of elastin chain structure. (**c**) Uncoiling of elastin molecules when stretched. When stretching force is relaxed, it will recoil spontaneously. (**d**) SEM image of elastin fibers

In Eq. 1, the crack (half) length is the initial machined notch of an impact test specimen:

$$K_{Ic} = \sigma_F(\alpha\pi c)^{1/2} \quad (4)$$

where α is a geometrical constant.

But, as illustrated, Eq. 3 is a measure of fracture toughness, a different measure than toughness. Fracture toughness, although related to toughness, describes the ability of a material containing a crack to resist fracture. It has units of MPa$\sqrt{m}$ in contrast to J/m^3 for toughness.

Elastin Structure and Function

Elastin is a natural polypeptide with distinct elastomeric properties and function. It serves an important function in arteries and larger elastic blood vessels such as the aorta. It also functions in lung tissue and especially skin where it allows skin to stretch and elastically relax to its initial position. It is involved in other internal organ soft tissue and elastic ligaments. *Fibrillin* also composes elastic fibers which usually contain up to ~75 % elastin. Major amino acid residues include glycine (32 %), alanine (21 %), proline (13 %), and valine, as well as lysine (see ▶ Fig. 1 in chapter "Structure and Function of Viruses and Bacteria") which cross-links the molecular chains when stretched. Elastin is often described as an amorphous polymer since it has no systematic or even semi-systematic structure or molecular array. As a long-chain polymer (Fig. 19a), it appears as represented schematically in Fig. 19b. When this cross-linked, polymer chain structure is stretched as illustrated in Fig. 19c, it forms linear strands but when the stretching force or stress is relaxed, the linear chain segments revert to the irregular (amorphous) polymer structure shown in Fig. 19b. The properties of elastin are similar to rubber or related polymeric materials as implicit in the SEM image of elastin fibers shown in Fig. 19d.

Elastin functions in concert with collagen which provides strength or firmness to related organs or organ tissue. In contrast to collagen fibers which have a nominal elastic (Young's) modulus of 1 GPa, elastin has a modulus of only 0.3–1.5 MPa. However, its strain exceeds 2 %, while collagen strains only one-twentieth of this value. Elastin is produced only up to the age of ~12 years in humans and then the body ceases to produce it, often resulting in a decrease with age. In addition, damage to elastin is incurred through UV radiation exposure and aging effects produce wrinkles in the skin.

References

Ballarini R, Heuer AH (2007) Secrets in the shell. Am Sci 95:422–429

Barani H, Montazer M, Samadi N, Toligat T (2012) In-situ synthesis of nano silver/lecithin on wool: enhancing nanoparticles diffusion. Colloids Surf B Biointerfaces 92:9–15

Chen P-Y, McKittrick J, Meyers MA (2012) Biological materials: emotional adaptations and bioinspired designs. Prog Mater Sci 57:1492–1704

Hosseini SH, Pairovi A (2005) Preparation of conducting fibers from cellulose and silk by polypyrrole coating. Iran Polym J 14(11):934–940

Johnston JH, Nisson T (2012) Nanogold and nanosilver composites with lignin-containing cellulose fibers. J Mater Sci 47(3):1103–1112

Keten S, Xu Z, Ihle B, Buehler MJ (2010) Nanoconfinement controls stiffness, strength and mechanical toughness of beta-sheet crystals in silk. Nature Mater 9:359–367

Liu G, Huang X, Wang Y, Zhang Y-Q, Wang X (2012) Thermal transport in single silkworm silks and the behavior under stretching. Soft Matter 8:9792–9799

Pauling L, Corey RB, Branson HR (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci 37(4):205–211

Perumalra R (2012) Effect of silver nanoparticles on wool fiber. ISRRS Chem Eng 2012, article 10842021, 4 pp

Saranathan V, Forster JD, Noh H, Liew S-F, Mochrie SGJ, Cao H, Dufresne ER, Prum RO (2012) Structure and optical function of amorphous photonic nanostructures from avian feather barbs: a comparative small angle X-ray scattering (SAXS) analysis of 230 bird species. J R Soc Interface 9(75):2563–2580

Seki Y (2009) Structure and mechanical behavior of bird beaks. PhD Dissertation, University of California San Diego, San Diego

Seki Y, Schneider MS, Meyers MA (2005) Structure and mechanical behavior of toucan beak. Acta Mater 52:2581–2596

Simmons AH, Michal CA, Jelinski LW (1996) Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. Science 271(5245):84–87

Singha K, Maity S, Singha M (2012) Spinning and applications of spider silk. Front Sci 2 (15):92–100

Termonia Y (1994) Molecular modeling of spider silk elasticity. Macromolecules 27 (25):7378–7381

Tong J, Ren L-Q, Chen BC (1995) Chemical constitution and abrasive wear behavior of pangolin scales. J Mater Sci Lett 14:1468–1470

Toni M, Dalle Valle L, Alibardi L (2007) Hard (beta) keratins in the epidermis of reptiles: composition, sequence and molecular organization. J Proteome Res 6:3377–3392

Varesano A, Tonin C (2008) Improving electrical performances of wool textiles: synthesis of conducting polypyrrole on the fiber surface. Text Res J 78(2):1110–1115

Wu X, Liu X-Y, Du N, Xu G-Q, Li B-W (2009) Molecular spring: from spider silk to silkworm silk. Appl Phys Lett 95:093703

Yao H, Dao M, Imholt T, Huang J, Wheeler K, Bonilla A, Suresh S, Ortiz C (2010) Protection mechanisms of the iron-plated armor of a deep-sea hydrothermal vent gastropod. Proc Natl Acad Sci 107(3):987–992

Comparison of Biological (Natural) Materials and Engineering Materials Properties

Contents

Abstract

A comparison of materials properties for natural (biological) materials and engineering materials is accomplished by reproducing Ashby plots or diagrams: Young's modulus versus density, toughness, fracture toughness, and strength, along with specific modulus (or stiffness) and specific strength.

Introduction

Although it may be somewhat apparent or even intuitive that biological materials have both similarities and differences in contrast to engineering materials, it may be useful to draw some visual comparisons prior to discussing biomimetics or biomimicry as related to innovations in materials structures or functions inspired by nature (Benyus 1997). A convenient format for such comparisons is embodied in the so-called Ashby maps, similar in concept to the deformation mechanism maps illustrated in ▶ Fig. 5a and b in chapter "Examples of Directional Crystal Structures: Gas Turbine Component Applications in Superalloys". These maps concentrate a large quantity of information into simple, graphic diagrams (Ashby 1992) which have been extended to biological materials by Wegst and Ashby (2004). The properties of interest and comparison have included density, Young's modulus, strength, and toughness. The interrelationships of these properties form useful comparisons especially in the context of design tools and the perusal of innovations related to biomimicry.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_29

Examples and Summary Tables

Figure 1a provides a graphic comparison of Young's modulus versus density for both natural and engineering composites, while in contrast Fig. 1b compares the fracture toughness and Young's modulus. Figure 2 shows, in contrast with Fig. 1a, the Young's modulus versus density emphasizing engineering materials, but

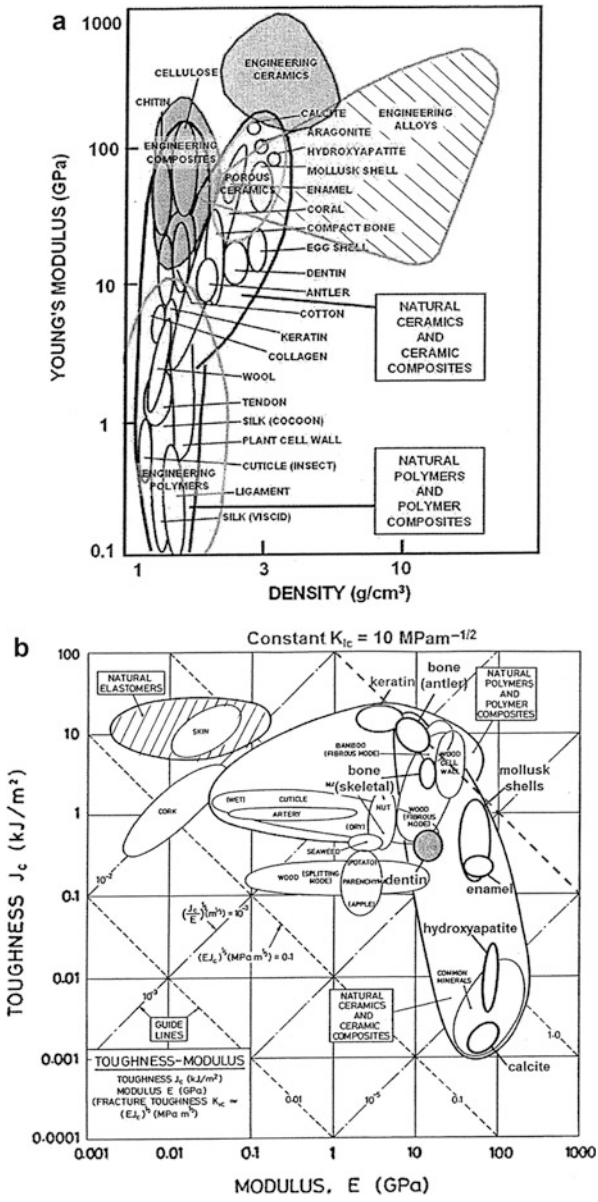


Fig. 1 Ashby plots for biological materials (**a**) Young's modulus versus density and (**b**) toughness versus Young's modulus (Adapted from Ashby 1992; Wegst and Ashby 2004)

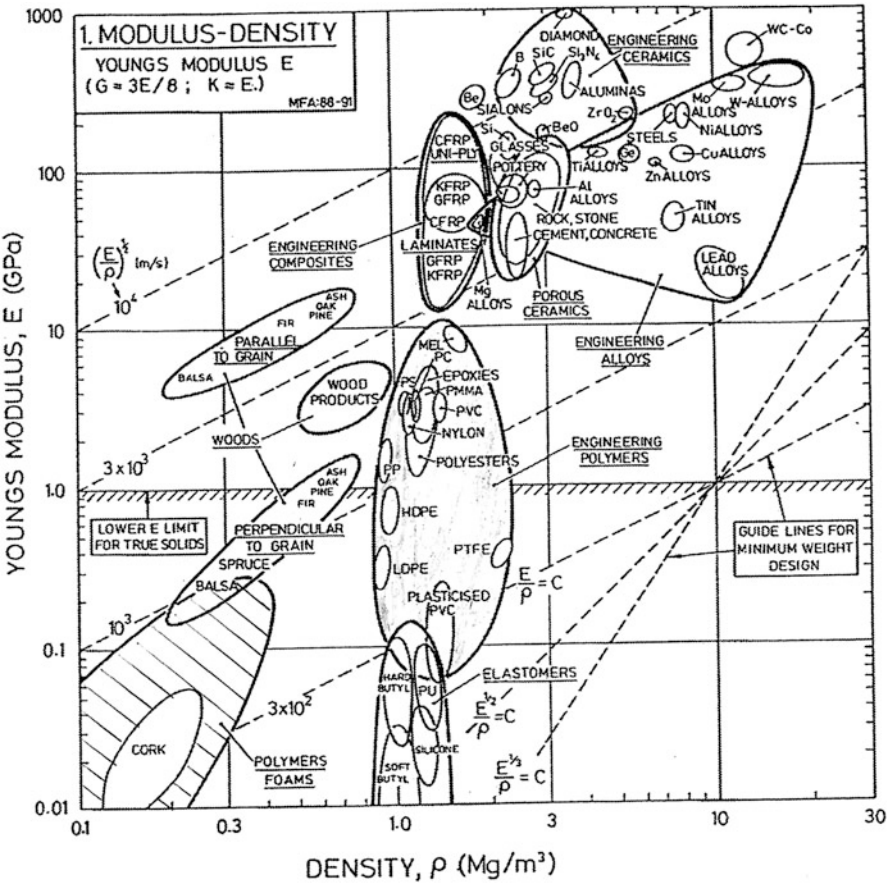


Fig. 2 Ashby plot of Young's modulus versus density for engineering materials (Adapted from Ashby 1992)

comparing a few natural materials such as wood. Various trends are shown by dotted straight lines which include the specific modulus having the notation E/ρ or $(E/\rho)^{1/2}$ and related forms. In Fig. 2 engineering materials are compared in more specific categories involving polymers, ceramics, and metals and alloys in contrast to more specific biological materials in Fig. 1a.

Figure 2 compares the fracture toughness and density more specific to engineering materials although wood provides some perspective in the context of natural

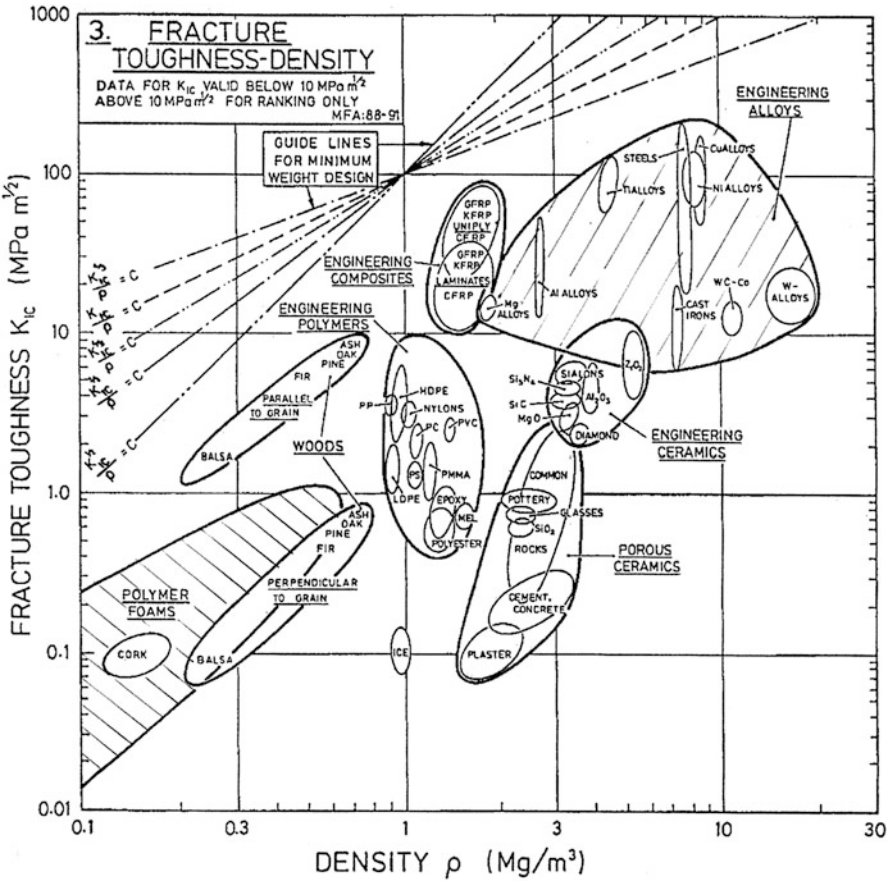


Fig. 3 Ashby plot of fracture toughness versus density for engineering materials (Adapted from Ashby 1992)

materials, which, as evident in Fig. 1a, have low densities, rarely exceeding 3 g/cm^3 (Mg/m^3). Figures 2 and 3 provide the contrast of fracture toughness versus Young's modulus for engineering materials in comparison to natural, biological materials in Fig. 1b. Here again the link is provided by wood in Fig. 4 although it must be recognized that the toughness in Fig. 1b is different from the fracture toughness in Fig. 4, although the Young's moduli generally extend over the same range representing roughly six orders of magnitude.

Figure 5 shows the fracture toughness versus strength mapped primarily for engineering materials although wood again provides the link to biological

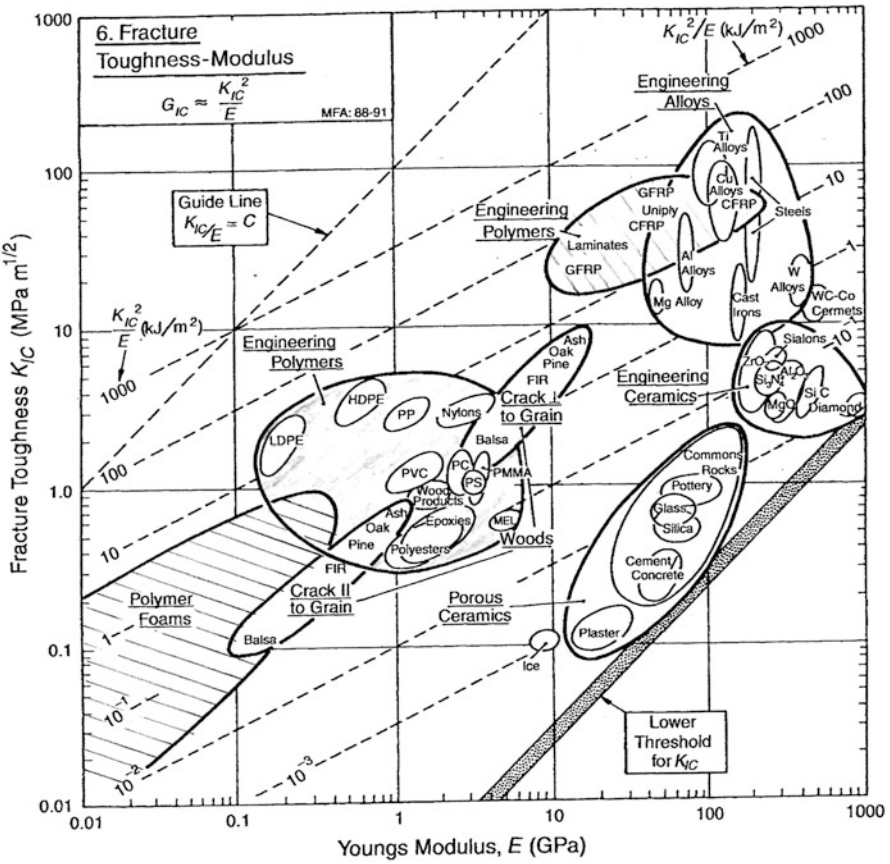


Fig. 4 Ashby plot of fracture toughness versus Young's modulus for engineering materials (Adapted from Ashby 1992)

materials. On comparing Figs. 4 and 5, it is observed that high modulus and high strength materials tend to have high fracture toughness. Indeed, Fig. 4 is generally represented by ▶ Eq. 3 in chapter "Structures and Properties of Keratin-Based and Related Biological Materials".

Figure 5 compares the specific modulus versus specific strength – E/ρ versus σ/ρ – for primarily biological materials in Fig. 6a with primarily engineering materials in Fig. 6b. These criteria, modulus-to-weight (density) or strength-to-weight (density), are often important design strategies for materials selection or improvement

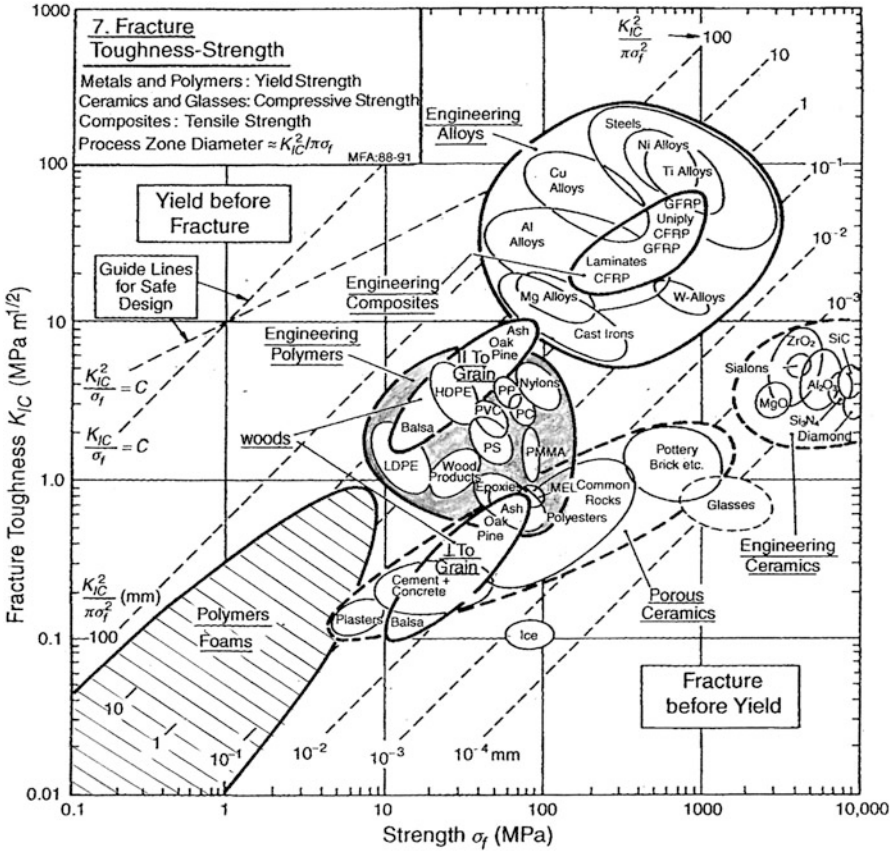


Fig. 5 Ashby plot of fracture toughness versus strength for engineering materials (Adapted from Ashby 1992)

in the context of materials performance issues, especially in automotive and aeronautical/aerospace applications, where strength and light weight are paramount.

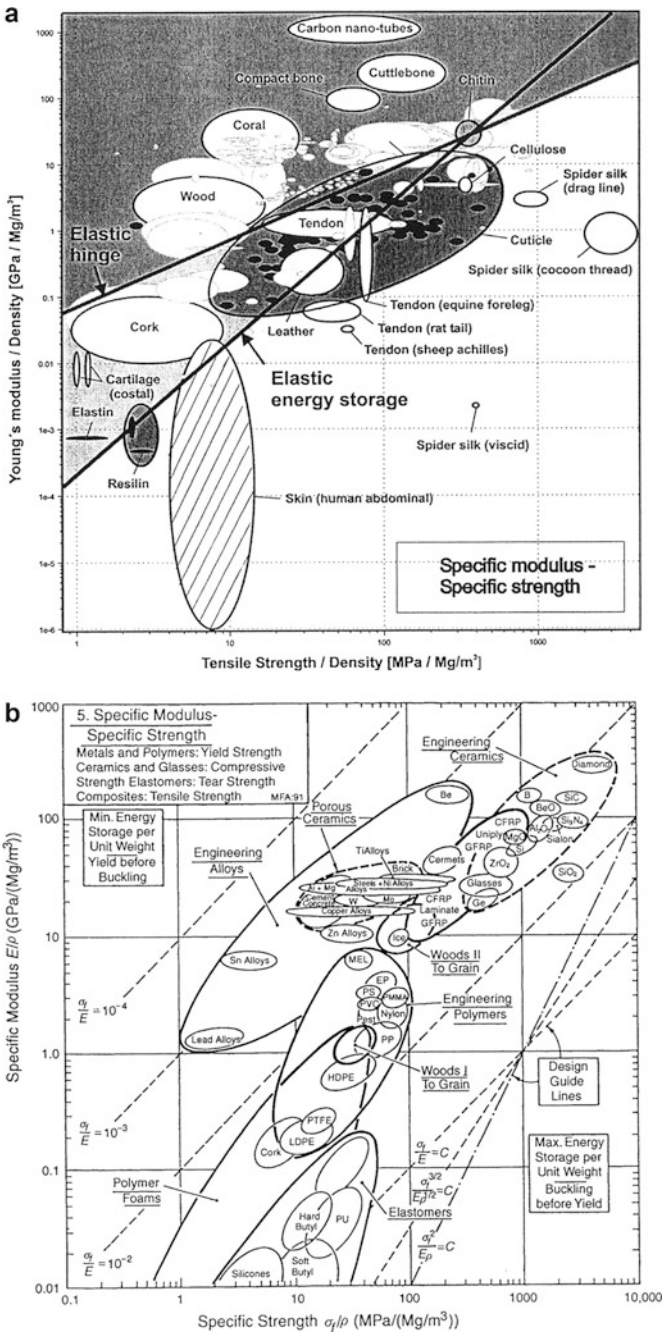


Fig. 6 Ashby plots of specific modulus versus specific strength. (**a**) For biological materials. (**b**) For engineering materials (From Ashby 1992; Wegst and Ashby 2004)

References

Ashby MF (1992) Materials selection in mechanical design, 4th edn. Butterworth-Heinemann, Oxford (2010)

Benyus JM (1997) Biomimicry: innovation inspired by nature. Morrow, New York

Wegst UGF, Ashby MF (2004) The mechanical efficiency of natural materials. Phil Mag 84:2167–2186

Part IX

Biomimetics

Biomimetics and Biologically Inspired Materials

Contents

Abstract

Nature provides a vast array of biological materials and materials systems which have inspired innovations in novel applications and in new materials developments. These have included natural (animal) armor, flight systems inspired of course by birds, fasteners and attachments, and an array of photonic structures. Microbes producing methane and keratin–rubber composites pose novel systems along with branched biological systems which include trees, lungs, circulatory structures, and the like, which are governed by fractal geometry. The concept of protein factories is revisited in this chapter where virus and bacterial systems can act as protein factories to produce a complex array of drugs and related organic materials and functional systems.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_30

Introduction: Bio-inspired Materials Design and Optimization Strategies

In the 1960s, a US Air Force researcher, Jack Steele, coined the word "bionics" in referring to copying nature or developing ideas from nature. More recently, the general concept of innovation inspired by nature has been interchangeably designated biomimicry or biomimetics, a term coined by Otto Steele in 1969. Biomimicry (from bios, meaning life, and mimesis, meaning to imitate) describes examining nature's successful developments and then imitating those designs, processes, and systems to solve problems with some benefits to humans Benyus 1997.

Biomimetics involve mimicking or synthesizing natural materials and biological systems or organisms to the practical design and fabrication of biomaterials or biologically inspired technologies (Bar-Cohen 2006). Organisms in nature have faced many of the challenges faced by humans or human society, but they meet such challenges by natural selection and evolution, a long-term process of optimization often requiring (or involving) millions of years. It is certainly advantageous to capitalize on nature's processes of optimization and adapt them to current human needs or challenges in materials and materials systems innovation.

Early humans observed nature and natural phenomena and replicated them or adapted them to their advantage in numerous says. These have included shelter and shelter architectures; weapons and defense, including armor, sensors, and alarm systems; and processes related to synthesis and manufacturing. Biological materials are apparently designed through a simultaneous optimization of structures and properties to achieve some functionality or to be multifunctional. From ▶ Figs. 1 to ▶ 6 of chapter "Comparison of Biological (Natural) Materials and Engineering Materials Properties" in particular, it can be observed that density (ρ) stiffness E, strength (σ), and toughness are central to optimizing structural support and flaw tolerance as related to degradation and failure.

As a process, biomimetics can involve the following:

- Identify materials issues or concerns and find natural materials or organisms exhibiting similar issues and observe their responses
- Embrace inspiration by observing and studying natural phenomena and natural materials systems: innovations inspired by nature
- Examine biomimetic simulations and explore their optimization in the context of natural phenomena and natural materials

Biologically Inspired (Bionic) Materials and Materials Systems in Antiquity: Armor and Flight

In ▶ Fig. 12e of chapter "Structures and Properties of Keratin-Based and Related Biological Materials," a pangolin coat was illustrated to represent an early example of an armored animal conceptually used for human protection. There are much older examples of the use of biological armors as human protective armor, and these

Fig. 1 Crocodile armor (helmet and body armor) on display in the British Museum, London (Courtesy of the British Museum, London)

most probably inspired the creation of similar armored structures using other materials such as metals, after metals, and metallurgy.

In the early Chinese Shang Dynasty (circa 1500 BC), nobles wore breastplates made from pieces of shell, especially turtle shells, tied together in an overlapping scale arrangement. Elite Egyptian bodyguards were also known to wear crocodile skins which were later replaced by overlapping plates of copper or bronze for Egyptian military officers and elite troops. These later metal armors were probably inspired by animal armors. A crocodile suite of armor illustrated in Fig. 1 was presented to the British Museum in 1846 which was recovered near Manfalut, Egypt. Since Roman legions occupying this territory circa AD 400 adopted the historical Egyptian culture, the suit in Fig. 1 is probably Roman rather than Egyptian.

Scale armor is probably the oldest type of body armor although other armor configurations of plates have been fabricated. Ancient Japanese armor used hexagonal-shaped leather tiles forming jackets or armor coats. The US Army tactical vehicles have also employed this configuration using hexagonal ceramic tiles coated with a polymer. Variances of this armor have included zirconia-toughened alumina ceramic coated with or encased in aluminum. Other ceramic cores have included Si, B_4C, TiB_2, etc., all forming relatively lightweight armor in

contrast to steel. In addition, these configurations cause projectile fragmentation and trajectory displacements in the ceramic section.

Reptile and fish scales have continued to inspire improved body armor. Bruet et al. (2008) have investigated the mechanistic origins of penetration resistance (or biting attack) of the scales of *Polypterus senegalus*, a "living fossil" fish. These were found to include multiple reinforcing composite layers that each undergo a unique deformation and crack arrest feature. Similar features have been investigated for the Arapaima, the 200 kg Amazon cod whose scales are largely impenetrable by piranha teeth (Lin et al. 2009). Yang et al. (2004) have also recently described bio-inspired flexible dermal armor using arapaima scale bionics along with a review of related natural armor biomimetics.

While flight of birds and insects has inspired many efforts to mimic flight over millennia, true flight bionics were only formally recorded by Leonardo da Vinci (circa AD 1496) who was unusually observant of nature. da Vinci's journals contain detailed studies of birds and several wing designs involving the so-called flying machines especially based on bat wing structure which he described as an impenetrable membrane. In the fourth century BC, the Chinese were known to use large kites to lift people to heights for navigation and communication over distances during battles, and the Montgolfier brothers tested the first hot air balloon around 1783 which was inspired by Roger Bacon's concept of air having substance and the prospects for raising people by lighter-than-air gas-filled balloons. It was not until the Wright brothers invented and built the first airplane (between 1901 and 1903) and demonstrated powered and sustained flight in December of 1903. They used unbleached muslin fabric covering the bottom surface of two wing sections resembling Chinese kites made of thin strips of ash wood. For some years, wood was the structural material of choice for aircraft until lightweight aluminum alloys became available. Airplane evolution, as illustrated in ▶ Fig. 7 in chapter "Examples of Man-Made Composite Structures", has involved fiber laminate composites along with other lightweight metals such as titanium and titanium alloys. Figure 2 illustrates the evolution of the airplane beginning with da Vinci's sketches (Fig. 2a), the 1903 Wright brothers' aircraft (Fig. 2b), and the more recent Eurofighter jet shown in Fig. 2c. The Eurofighter in Fig. 2c is composed of ~70 % carbon fiber composite shell (wing skins) and structural members, and a Li–Al alloy on the leading edge of the wings. Fiberglass-reinforced polymeric composites compose ~12 % of the structure with only 15 % metal components, mostly Ti, and including the Li–Al alloy wing sections. It is powered by two turbofan engines which produce speeds up to 2,125 km/h, more than twice the speed of sound.

Biological Attachment and Adhesion: Velcro and Animal-Inspired Dry Adhesives

In addition to the mimicking of animals for armor configurations and efforts to fly as observed in birds and insects as illustrated above, there are other novel biomimetic examples which provide rather simple and rather complex concepts. The first

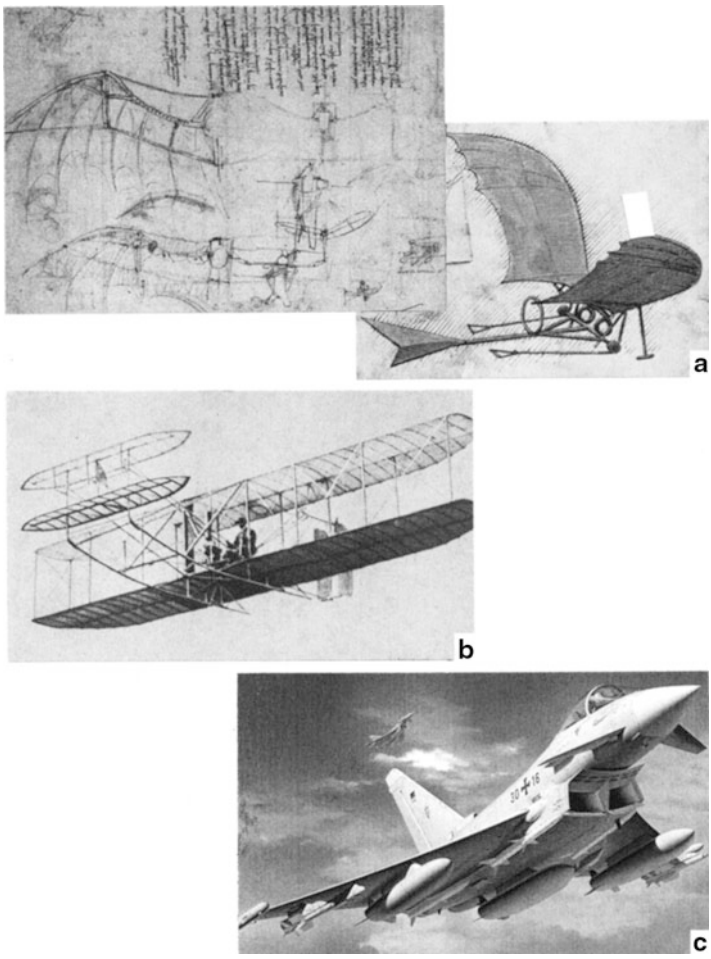


Fig. 2 Aircraft evolution. (**a**) da Vinci drawings. (**b**) 1903 Wright brothers aircraft. (**c**) 2011 Eurofighter aircraft (Google images)

involves the development of Velcro which evolved by observations of natural burrs by George de Mestral, a Swiss engineer, on the fur of his dog around 1940. Upon examining the burrs in a light microscope, de Mestral was struck with the prospects of creating a fabric fastener based on the hooks he observed at the end of the burrs. However, it was not until more than 25 years later that Velcro was developed as a commercial product.

A more timely and correspondingly more complex attachment phenomenon was observed for the gecko as described in connection with ▶ Fig. 1 in chapter "Chemical Forces: Nanoparticles" which illustrates the importance of the hydrogen or van der Waals force created by closely spaced arrays of keratinous setae having spatulae at their ends covering the gecko feet. Yurdumankan et al. (2005) used

multiwall carbon nanotube arrays to mimic these arrays of gecko foot setae while Geim et al. (2003) utilized microfabricated arrays to achieve the same adhesion phenomenon. Similar arrays of 400 nm wide silicone pillars were developed by Lee et al. (2007) while Yu et al. (2001) used micrometer-size polymer arrays to create dry adhesion for robotic applications. The intermolecular potential created by setae and synthetic setae arrays is given by

$$\phi = -\frac{\rho_1\, \rho_2\, \pi^2 \alpha\, R}{6D} \tag{1}$$

where ρ_1 and ρ_2 are the number of contacts of two surfaces, R is the radius of each contact, D is the distance between the two surfaces, and α is a geometrical constant. For the tokay gecko foot in ▶ Fig. 24 of chapter "Electromagnetic Fundamentals," there are ~14,400 setae/mm^2 or ~3 $\times$ 10^6 in the two front feet which can produce a force ~40 times the gecko weight. Similar analyses by Arzt et al. (2003) predict the spatula diameter for the van der Waals force – related adhesion will decrease with the body mass to the $-1/3$ power ($m^{-1/3}$). Correspondingly, for smaller mass organisms such as beetles with a mass near 0.2 g, the spatula diameter will be ~3 μm. It is also notable that the gecko adhesive is self-cleaning, and gecko tape and related next-generation adhesives modeled after gecko toes have been discussed by Autumn and Grovich (2008).

Mussels attach to rocks through arrays of byssal threads having special seta properties involving uniaxially aligned nanofibers which are still not well understood, although a combination of mechanical interlocking with non-covalent interactions, including hydrogen bonding and van der Waals bonding, comes into play. Underwater adhesion of the abalone foot involves similar arrays of setae (Lin et al. 2009).

Plants can involve similar, complex attachment strategies. The English ivy plant involves microfibril orientation in the root hairs and other mechanisms such as substructural changes at the subcellular level that work together to allow the plant's attachment to various climbing surfaces (Meizer et al. 2010). The English ivy root hairs also secrete a nanocomposite adhesive containing spherical nanoparticles 60–85 nm in diameter. These organic nanoparticles show promise in biomedical and cosmetic applications as a safer alternative to metal oxide nanoparticles such as Fe_2O_3 or TiO_2 (Burns et al. 2012).

The principal feature for biomimetic considerations for dry and wet adhesives generally involve attachment mechanisms that minimize or eliminate toxic glues, fasteners, clips, tapes, and related issues.

Light-Emitting (Bioluminescent) Organisms

Light-emitting (bioluminescent) organisms which convert chemical energy into light include an amazing diversity of organisms including insects, most notably the firefly, a variety of jellyfish, and certain bacteria, among others. Some organisms emit light by absorbing specific wavelengths while others emit light by absorbing energy of light. Bioluminescence evolves from chemical reactions as is

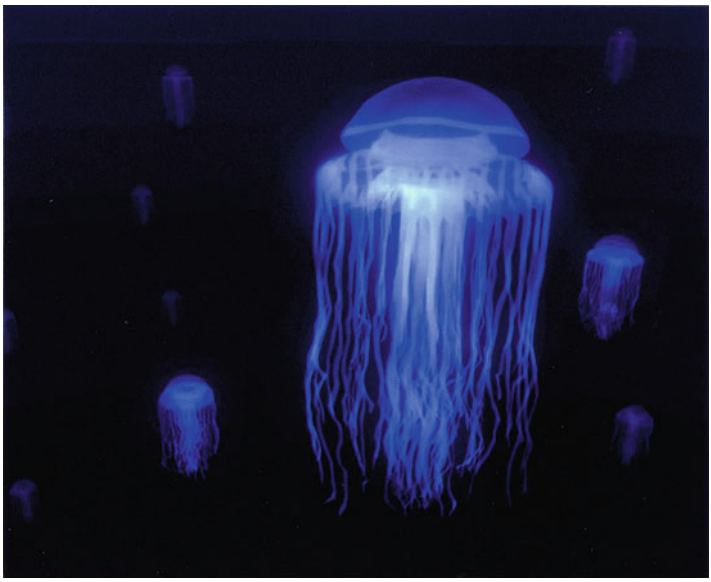


Fig. 3 Bioluminscent jellyfish (Photograph courtesy of Phillip Colla (Oceanlight.com)). The majority of bioluminescent organisms are ocean dwellers

the case with fireflies and noctiluca, sea-dwelling species that exhibit bioluminescence as illustrated in Fig. 3. Although proteins are involved in this process, they do not emit light directly. Light emission from the North American firefly (*Photinus pyralis*), a yellow-green (557 nm) light, is believed to be the most efficient bioluminescence system known. The chemical reaction catalyzed by firefly luciferase occurs in two steps as follows:

$$\begin{aligned} &1.\quad \textbf{Luciferin} + \textbf{ATP} \rightarrow \textbf{luciferyl adenylate}\ (\textbf{LA}) + \textbf{PPi} \\ &2.\quad \textbf{LA} + \textbf{O}_2 \rightarrow \textbf{oxyluciferin} + \textbf{AMP} + \underline{\textbf{LIGHT}} \end{aligned} \tag{2}$$

where ATP is the universal biochemical energy source, adenosine triphosphate; PPi is inorganic pyrophosphate; and AMP is adenosine monophosphate. Firefly luciferin, ATP, LA, PPi, oxyluciferin, and AMP are illustrated in Fig. 4 which is a graphic for Eq. 2. Light emission occurs in Eq. 2 and Fig. 4 because the oxyluciferin is formed in an electronically excited state and a photon is released when the structure returns to the ground state by electron recombination or fluorescence. Bioluminescence is therefore an O_2 consumption pathway present in luminous organisms.

Luciferase can be produced by genetic engineering and the luciferase genes can be synthesized and inserted into organisms or transferred into cells. These include silkworms and mice which have been engineered to produce the protein (Contagal and Bachmann 2002). Recently, a single human cell and jellyfish protein (luciferase) (Fig. 3) were observed to create a single-cell laser (Gather and Yun 2011).

Kim et al. (2012) have investigated high-transmission microstructures in a firefly lantern and its biomimicry for efficient light-emitting diode (LED) illumination.

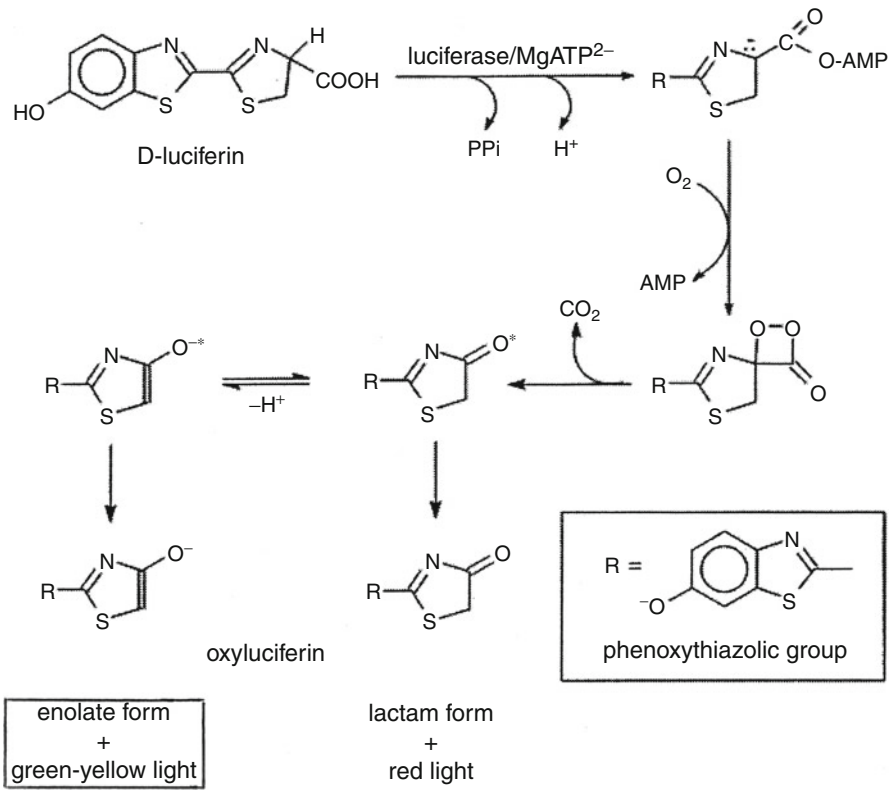


Fig. 4 Bioluminescence chemical reaction graphic characteristic for the North American firefly. Two light emission (color) pathways are illustrated. Visible light emission results from the rapid loss of energy of the excited state oxyluciferin molecule via a fluorescence pathway

In this respect, cuticular nanostructures found in the firefly lantern have been observed to manage light for structural color, optical index matching, or light polarization within an ultrathin natural scale. These nanostructures are involved mainly in the management of incoming light and have inspired recent display and imaging applications (Kim et al. 2012). The bioluminescent firefly lantern creates highly efficient light signals in sexual communication (Lewis and Cratsley 2008). While the light-producing reactions have been well understood as implicit in Fig. 4, the cuticular nanostructures are less well known but, as illustrated in Fig. 5, have inspired an LED lens distinct from a smooth surface lens as illustrated in Fig. 6. This cuticular-mimicked lens substantially increased light transmission in the visible portion of the spectrum comparable to conventional antireflection coatings. This can offer new opportunities for increased light extraction efficiency of high-power LED systems.

Alam et al. (2012) have also recently demonstrated the chemical reactions taking place between a light-emitting substance and its related enzyme: the luciferin-to-luciferase reaction (Fig. 4), respectively, could provide a natural, electricity-free

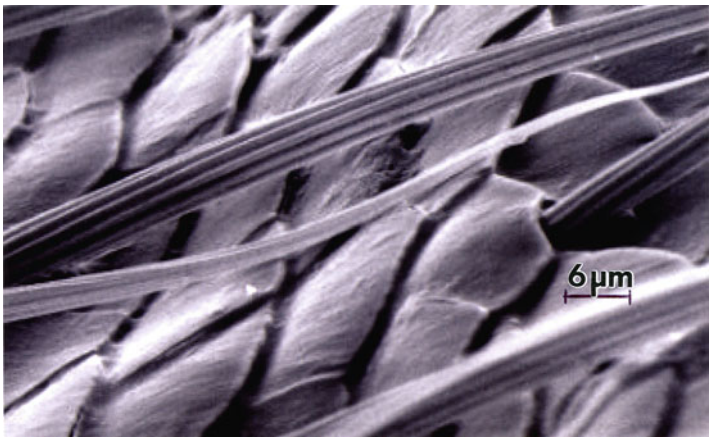


Fig. 5 SEM image of misfit or serrated scales on the photuris firefly abdomen (or lantern) which enhance the glow emitted from it by minimizing internal reflections of light and maximizing the light intensity emitted (After Bay et al. 2013)

glow regime for ambient lighting and other uses. They demonstrated this phenomenon by doping Cd-based semiconductor nanorods with luciferase and found the luciferin-to-luciferase reaction was 20–30 times more efficient than other attempts at producing bright light.

Bioluminescent devices for zero-electricity lighting through the manipulation of bioluminescent populations of microorganisms have been investigated using *Vibrio fischeri*, a bacteria that naturally glows in the dark, and *Pyrocystis fusiformis*, an algae species that glows when excited by movement (Armstrong and Spiller 2011). In addition, experiments with genes to allow bioluminescence to be implanted in different organisms, even trees to produce "trees that glow," could replace streetlights! This concept is reminiscent of the "glowing tree of Souls" in the movie, Avatar, which, according to the movie, had the ability to directly connect the nervous systems of all living things.

Photonic Structures and Related Functional Systems

Although structural color and photonic crystals were presented briefly in chapter "► Electromagnetic Color and Color in Materials" (Fig. 14) and applications of photonic materials will be reviewed in Part XV: Innovations in Magnetic, Electronic, Optoelectronic and Photonic Materials, it may provide some continuity to present a short overview of aspects of bio-inspired optics. To begin, the reader is referred to recent books outlining biomimetics in photonics (Karthaus 2013) as well as optical biomimetics (Large 2013). As illustrated in ► Fig. 14 of chapter "Electromagnetic Color and Color in Materials" and Figs. 5 and 6, the most finely tuned, rapidly responsive, and precisely directed optical systems can be found in the surfaces of living organisms.

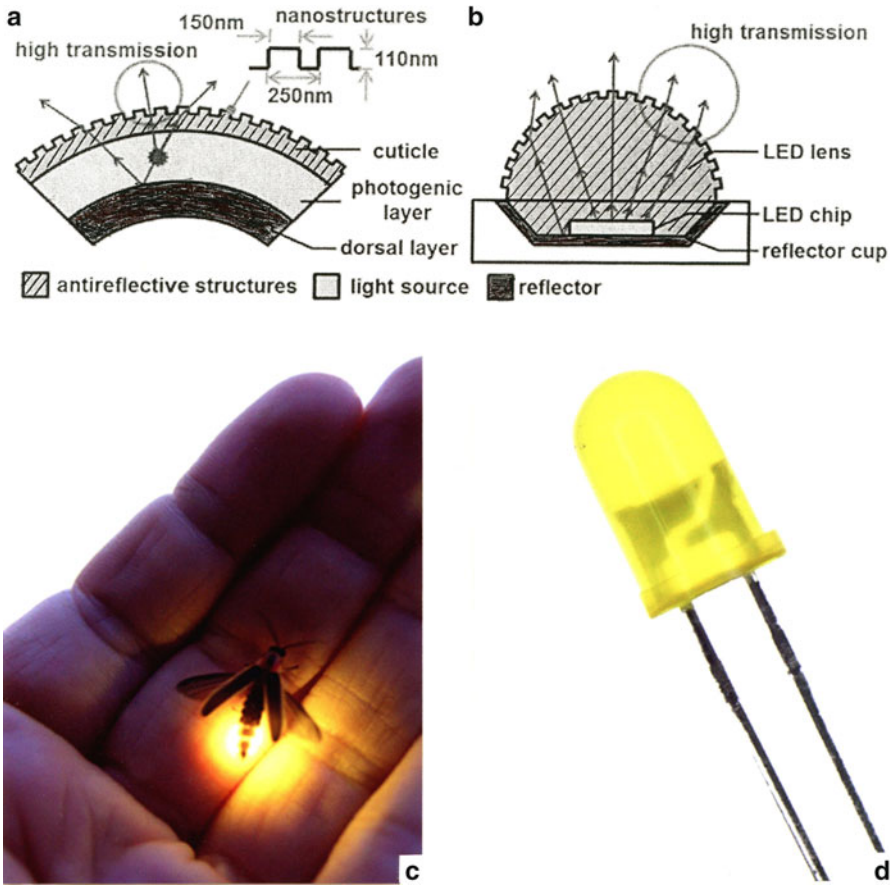


Fig. 6 Mimicking the cuticular nanostructures of the firefly lantern on the lens of a high-power LED for light extraction efficiency. (**a**) Firefly lantern schematic. (**b**) Biomimetic LED structure (Adapted from Kim et al. 2012). (**c**, **d**) illustrate firefly and LED light emission, respectively

Figure 7 shows a stereomicrograph of a microleaf beetle (*Chrysolina fastuosa*) on the head of a pin exhibiting iridescent colors typical of many beetle and other insect surfaces. These structural colors are generated when surface structures have been preserved in fossil beetles which exhibit vivid colors similar to Fig. 7 generated by chitin multilayer reflector structures in the epicuticle (McNamara et al. 2012; Seago et al. 2009). These create interference colors. Other cuticular structures contain 3D arrays which function as opal-like photonic crystals (► Fig. 14g of chapter "Electromagnetic Color and Color in Materials") and create diffraction colors. In addition, diffraction grating-like structures similar to the butterfly wing (► Fig. 14e of chapter "Electromagnetic Color and Color in Materials") also create color. In some multilayer cuticular chitin stack compositions, the stock is composed of layered pairs with different refractive indices and reflects all

Fig. 7 Micro-leaf beetle (*Chrysolina fastuosa*) on the head of a pin (Photograph by Klaus Bolte, Ontario, Canada)

types of light. These also are characterized by Bouligand structures where a twist of layer directions reflects circular polarized light. In these structures the color is created by diffraction. Bouligand structures or the so-called twisted plywood structures have also been discussed previously in the context of natural composite strengthening mechanisms in ▶ Fig. 14a and ▶ 15b of chapter "Structures and Properties of Keratin-Based and Related Biological Materials". Figure 8 illustrates examples of these structural color regimes. Helicoidal Bouligand structures also provide strength to beetle shells and other shell structures (Hepburn and Ball 1973).

An interesting feature of Fig. 8 is that these structural features can be mimicked using focused ion beam chemical-vapor deposition (FIB-CVD), direct laser writing, or other 3D printing technologies. Consequently, iridescent surface colors or imagery can be computer generated as permanent surface nanostructures which, if not degraded, can maintain these color features indefinitely.

Figure 8h, in addition to illustrating the photonic gratings in Fig. 8b, c in particular, also illustrates other functions of these and similar nanostructural arrays. These include wettability as it relates to hydrophilic and hydrophobic phenomena represented by the beetle shell features in Fig. 8d. Initially, water vapor (or mist) can collect on hydrophilic surfaces by heterogeneous nucleation controlled by the contact angle (Ω_c) between the nucleating water droplet and the surface:

$$\Omega_c = \cos^{-1}\left[(\gamma_{SV} - \gamma_{SL})/\gamma_{LV}\right] \tag{3}$$

which is Young's relation, where γ_{SV}, γ_{SL}, and γ_{LV} are the solid–vapor (SV) (or solid substrate) interfacial free energy, solid–liquid (SL) interfacial free energy, and liquid–vapor (LV) interfacial free energy, the surface tension of the water droplet. For a droplet forming homogeneously in the vapor phase or forming a contact angle of 180° with the substrate, the droplet becomes completely hydrophobic or super hydrophobic: no wetting of the surface. These conditions are illustrated generally in Fig. 9.

As implicit in Figs. 8h and 9, wettability incorporates both the topographical and chemical (or chemical potential) components of a surface, and these are also related

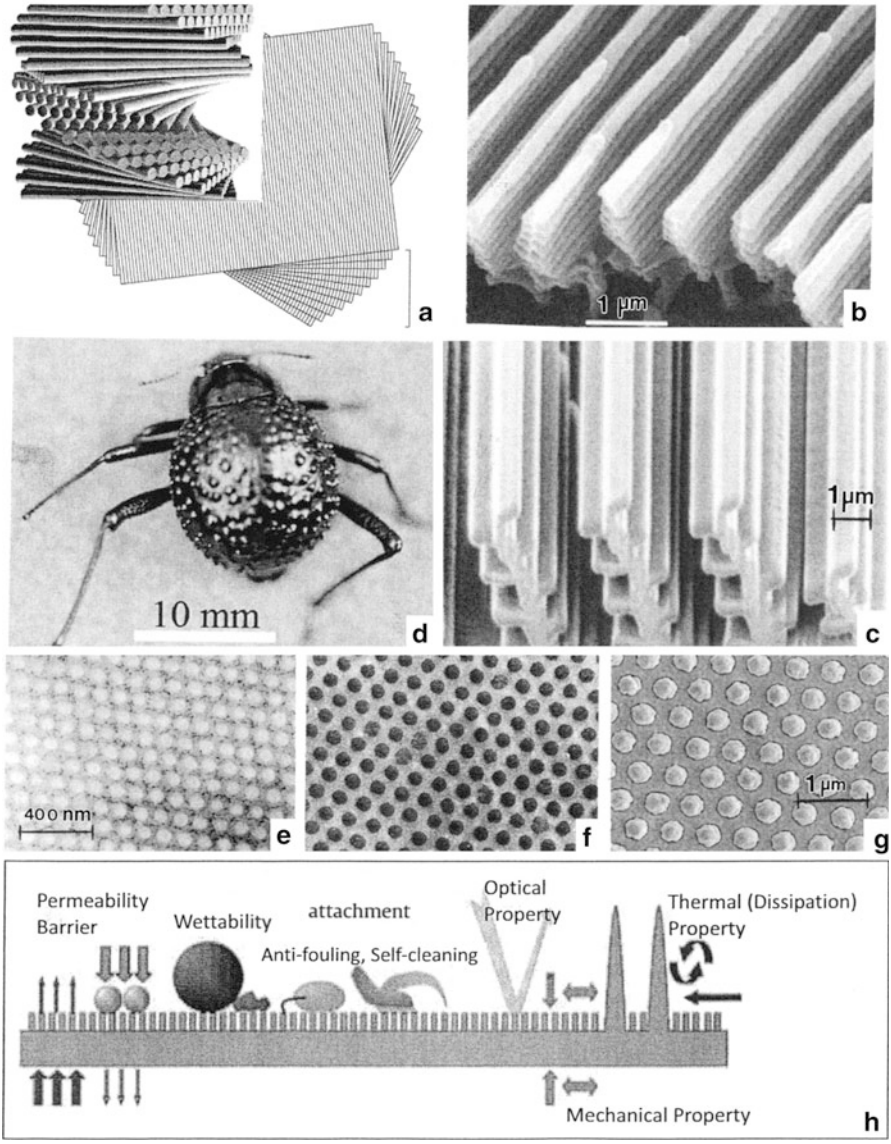


Fig. 8 Biological structures originating structural color and related multifunctionality. (**a**) Bouligand structures observed at 90°. Chitin layers rotated to create birefringent chiral nanofiber arrays (▶ Fig. 15 in chapter "Structures and Properties of Keratin-Based and Related Biological Materials") on shells of some beetles produce interference colors. These can include layered pairs of different refractive indices. (**b**) SEM image of *Morpho didius* ground-scale structure – a butterfly species with large blue wings which can change to violet with a viewing angle change. Ridge spacing produces diffraction (grating) color: ~450 nm (*blue*) (Kinoshita et al. 2008). (**c**) FIB-CVD carbon replica of butterfly wing structure in (**b**) (Adapted from Watanabe et al. 2005). (**d**) Wax-free, chitin-based hydrophilic bumps on elytra (forewing) of the desert beetle *Stenocara*

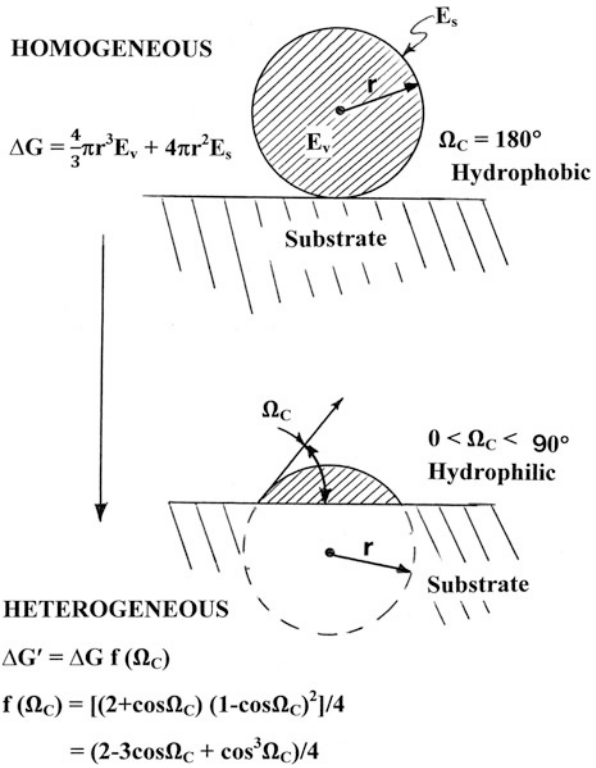


Fig. 9 Fundamental hydrophilic/hydrophobic concepts related to homogeneous and heterogeneous nucleation: contact angle phenomena. The wetting is hydrophobic for $90° < \Omega_c = 180°$ and superhydrophobic for $\Omega_c \cong 180$. Hydrophilic surfaces are characterized by $0 < \Omega_c < 90°$ and superhydrophilic where $\Omega_c \approx 0°$. The organism sensing approaching predators. In some hairs containing liquid-filled canals, irradiation by infrared light (~3 μm wavelength) causes thermal expansion of the liquid and expansion of the sensory hair, stimulating a mechanical signal to be sensed by the nervous system of the organism (Shcmitz et al. 1997). Other multiple arrays of spherical sensory cells called sensillum arranged on the backs of complex eyes (Fig. 10) perform similar sensory functions

to surface adhesion and friction. Many insects, especially butterflies, exhibit non-wetting or superhydrophobic surface features which are a fundamental survival issue in affording resistance to wetting by rain or other hydrated surfaces they may encounter. Topographies which systematically increase the air–water interface and

Fig. 8 (continued) sp. capture moisture beads which coalesce into larger drops which run down the beetle superhydrophobic, wax-coated shell surface into its mouth. There are 400,000 beetle (Coleoptera) species, the most diverse insect order (From Parker and Lawrence 2001). (**e**, **f**) show color-producing collagen arrays (nano-bumps and holes) from caruncle tissue of light blue and green birds endemic to Madagascar (*Philepitta castanca*) (From Prum et al. 1999). (**g**) Gold nano-bumps produced by photolithography. (**h**) Biological surface features producing multifunctionality (Adapted from Shimomura 2010)

minimize the solid–liquid contact area will lead to higher (more hydrophobic) contact angles, Ω_c. Correspondingly, the reverse conditions will lead to superhydrophilicity (Martines et al. 2005). As water droplets roll off the wing or shell surface of insects, dirt and other particulate contaminants are removed, creating an antifouling and self-cleaning feature. In addition, surface features exhibiting high contact angles also exhibit low-particle adhesion forces as well (Ralston and Swain 2011). The cleaning effect resulting from surface structure is often referred to as the Lotus effect since arrays of bumps several microns in size on the lotus leaf surface (similar to the beetle bumps shown in Fig. 8d) and protrusions of waxlike microcrystals form arrays, even pleats on these bumps. The resulting *fractal* undulating structure provides the lotus leaf surface a superhydrophobic property (Tsujii 2009).

There are numerous examples of scales or scale structure exhibiting low friction features. These involve desert rattlesnakes and lizards, especially the *Scincus scincus* lizard which inhabits the deserts in North Africa and Southwest Asia. Also called a "sand fish," this lizard dives into the sand and moves below the surface in a swimming-like fashion. Its skin scales have a smaller coefficient of friction than polished steel or Teflon and are characterized by surface arrays having heights of several hundred nanometers, spaced several microns (Baumgartner et al. 2008; Berthe et al. 2009). There are currently several commercial products which coat surface to produce superhydrophobic phenomena as well as fabrics woven with silica-coated spheres to produce nanostructures conducive to superhydrophobic behavior.

There are also of course larger insect structures which have inspired biomimetic materials and devices. These involve larger-scale (micron-size) features such as compound eye structure and hair structure. These are illustrated in Figs. 10 and 11, respectively. Even though these are larger structural arrays, they are composed of the same hierarchy of molecules, fibrils, fibers, and composites characteristic of all biological systems.

Compound eye structures as illustrated characteristically in Fig. 10 have inspired numerous biomimetic applications which take advantage of the their non-reflective surface structures which exhibit periodic or repetitive patterns similar to those in Fig. 8. The moth eyes have been a particular model for enhancing solar cell efficiency (Zhao et al. 1999) as well as solar-collector efficiency and large-area antireflective film on display devices (Wilson and Hutley 1982).

Figure 11 illustrates insect and other animal hair structures as sensors and the prospects for biomimetic sensor and sensor materials development. Some hairs function as mechanosensitive sensors where their movement in airflow or change in airflow can allow the organism to sense approaching predators. In some hairs containing liquid-filled canals, irradiation by infrared light ($\sim$ 3 microns wavelength) causes thermal expansion of the liquid and expansion of the sensory hair, stimulating a mechanical signal to be sensed by the nervous system of the organism (Schmitz et al. 1997). Other multiple arrays of spherical sensory cells called sensillum arranged on the backs of complex eyes (Fig. 10) perform similar sensory functions.

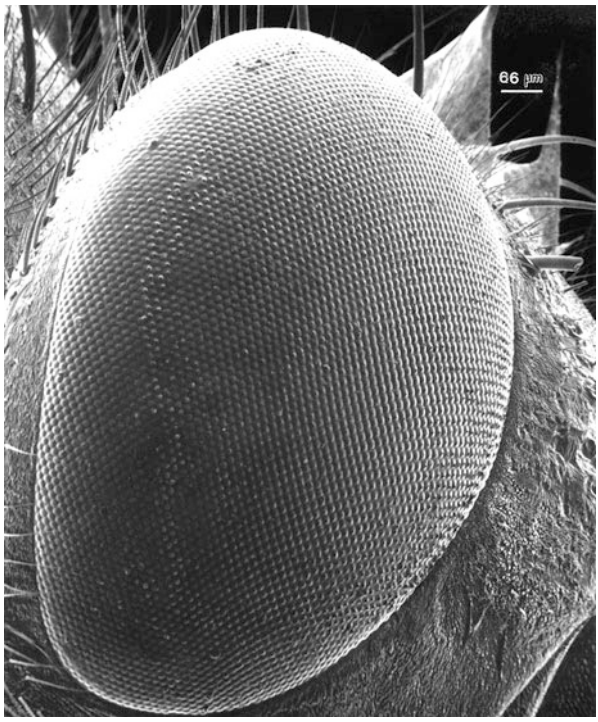


Fig. 10 SEM image of insect eye

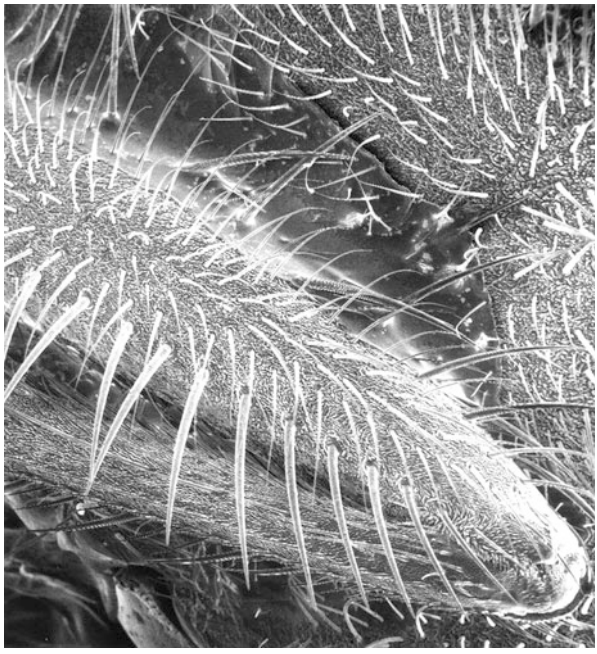


Fig. 11 Insect leg and body hairs observed in the scanning electron microscope. Magnification is the same as shown in Fig. 10

It is clear in retrospect, and on considering the implications of Fig. 8, especially Fig. 8h, that multifunctionality is certainly related to structural hierarchy. Replicating these structural hierarchies will be the key to the development of novel biomimetic materials and systems in the future, especially where these can be fabricated by 3D printing technologies to be described in Part XI: Advanced Materials Processing and Fabrication: 3D Printing and Additive Manufacturing Technologies following.

Keratin–Rubber Composite and Other Examples of Biocomposites

An example of the use of natural, biological material in the creation of a rubber-based or tire composite (► Fig. 9 in chapter "Examples of Man-Made Composite Structures") is described in a recent patent application (Hergenrother et al. 2012). In this application, β-keratin in the form of avian feather quill and fibers (including chickens and other fowl) representing several billion pounds of discarded feathers in the US poultry industry is added or pre-dispersed into synthetic or natural rubbers in place of other fillers such as black carbon. A coupling agent is used to chemically bind the keratin in the form of feather meal or ground feathers to the rubber. The keratin ranges from ~1 to 500 μm in diameter and 0.01–50 mm in length. In the context of the product discarded in the poultry industry, this application would represent a significant example of by-product synergism: the use of waste material in the creation of new, useful materials or products. Similar products have included plant waste composites and various wood products from wood chips and shavings in polymeric binders or matrices.

Of course there are hundreds if not thousands of examples of biological materials as waste fibers used in composite development. A literature review of nonwood plant fibers for building materials, for example, lists 1,165 references worldwide (Youngquist et al. 1994). While not specifically involving biomimetics, such examples illustrate the by-product synergism of biological waste materials.

Branched Systems and Fractal Structures

Biological systems are predominantly complex, irregular, and correspondingly nonlinear. Such irregularities are not quantifiable by classical Euclidean geometry. To overcome these geometry limitations, Mandelbrot (1982, 2012) formalized the concept of fractal dimension and introduced the term "fractal geometry." While Euclidean geometry deals with objects in integer dimensions, fractal geometry deals with non-integer dimensions, referred to as the fractal dimension. Fractal geometry is fundamentally self-similar. Self-similar growth is ubiquitous to nature and is governed by universal dynamical laws which are independent of the exact details of the dynamical system, e.g., chemistry, physical properties, etc., a system which evolves with time. The human body is a good example of fractal geometry.

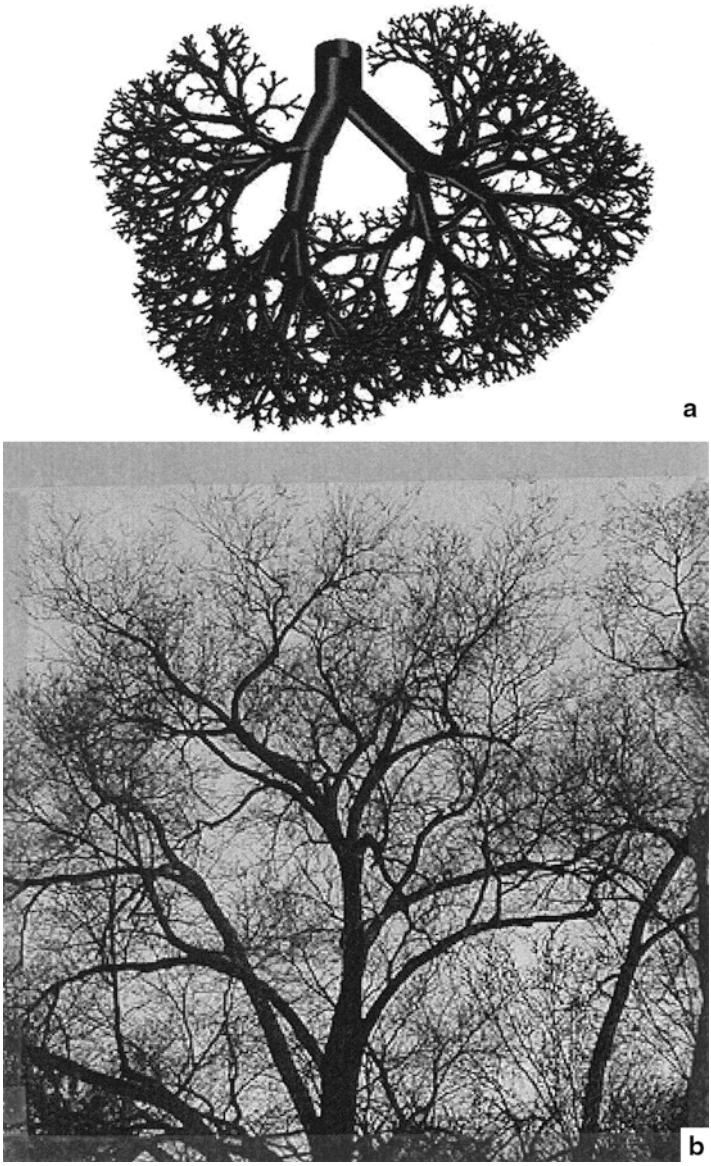


Fig. 12 Fractal branching. (**a**) Human lung. (**b**) Tree branching

Particular examples include the bronchial tree branching illustrated in Fig. 12a in contrast to regular tree branching in Fig. 12b. These illustrate the concept of fractal dimension:

$$M^D = N \tag{4}$$

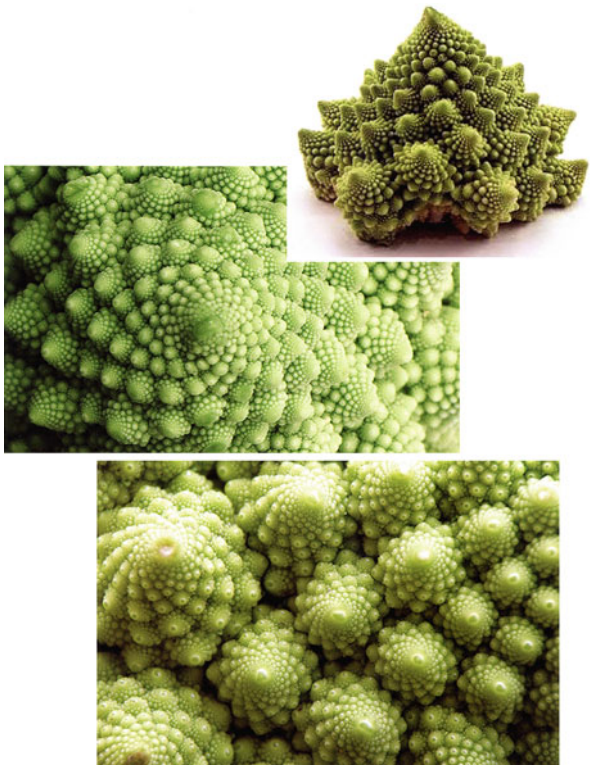


Fig. 13 Fractal patterns in the Romanesco broccoli showing natural self-similarity

where M is the magnification, D is the fractal dimension, and N is the number of identical shapes, where from Eq. 4

$$D = \log N / \log M \tag{5}$$

The basic fractal concept embodies patterns which when taken as the smallest part appears similar if not identical to its full structure. This feature is implicit in the branching patterns in Fig. 12 which are characteristic of a wide variety of hierarchical systems such as dendritic and microdendritic growth of many materials, particularly metals and alloys. Many fractal patterns illustrate an evolving self-similar or hierarchical structure within a sea of randomness. In this sense, the cauliflower or cauliflower-like fractal structures, such as the Romanesco broccoli, are among the most fascinating natural fractal forms, as illustrated in Fig. 13. These forms also illustrate regular sequences as shown by the directions of the arrows in Fig. 13 in the evolution of self-similar shapes, sizes, and relative positions of the fractal components. In this sense, we can rearrange Eq. 5 in the form

$$D = [\log(Nn + 1/Nn)] / [\log(rn/(rn + 1)] \tag{6}$$

where Nn is the fractal set or number of fragments with the linear dimension defined as rn.

It is also noted that the broccoli pattern in Fig. 13 also represents a Mandelbrot set represented by an equation of the form

$$Z = Z^2 + C \tag{7}$$

where both Z and C are complex numbers and the Mandelbrot set is a set of all complex C such that $Z = Z^2 + C$ does not diverge. C for a set is a constant number meaning that it does not change as the set evolves. If we start at $Z = 0$, it changes with each iteration which creates a new Z equal to the old Z squared plus the constant C.

While Figs. 12 and 13 illustrate fractals and fractal geometry in biology and nature, there are applications of fractals in biology and medicine related to understanding practical phenomena. These include brain and lung structure (Fig. 12a), DNA sequencing, and long-range anticorrelations in heartbeat regulation or rhythm (Havlin et al. 1995). Fractal shapes and fractal analysis have also been applied to the identification of tumors in brain MR images and other histological images in medical diagnostics (Klonowski et al. 2010). Fractal antennas, which use self-similar design to maximize material structures which receive or transmit electromagnetic radiation, produce compact, multiband, or wideband antennas used in cell phone and other microwave applications.

Because dendritic structures occur as a consequence of heat flow and because similar natural structures such as blood vessels and systems are represented by the complex branching shown in Fig. 12, similar fractal branching systems, even complex 3D systems, can be considered as efficient heat transfer or heat exchange devices (Rayneau-Kirkhope et al. 2012). These can be fabricated from a variety of materials and especially using 3D printing concepts. Such fractal systems might be especially applicable in cooling small devices and electronic circuits, including LED components.

These designs can be printed on virtually any surface or printed circuit configuration. Fractal branching as expressed by trees as illustrated in Fig. 12b might also be exploited in some new form of solar cell design. Ultralight fractal structures representing hierarchical structures have also been described as a design route for lightweight multifunctional structures (Rayneau-Kirkhope et al. 2012). Such intricate structures (Fig. 14) can be fabricated by 3D printing or related additive manufacturing concepts to be discussed in the next chapter (Part VIII: Biological Materials).

Biomimetic Design of Composite Interfaces

Although we have described composites and composite systems have been described in Part VIII: Biological Materials and including biological materials systems in Part VIII: Biological Materials and in this chapter, in Part IX: Biomimetics, it may be prudent to look at the issue of interfaces. In retrospect, it can be appreciated that man-made composites exhibit poor compatibility at the composite interface. However, in biological systems, interfaces exhibit outstanding stability,

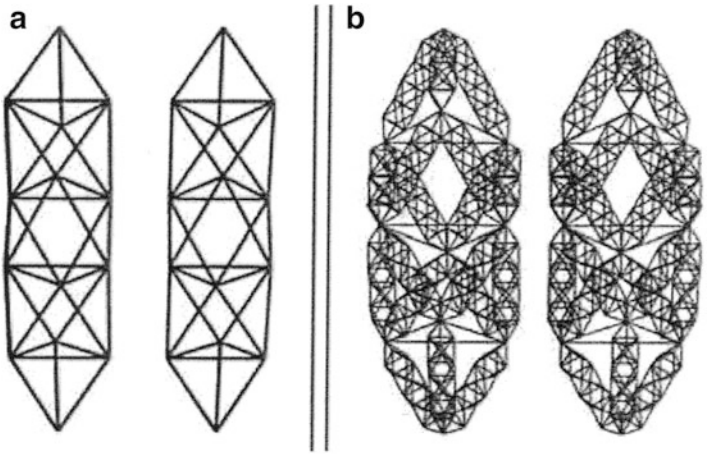


Fig. 14 Load-bearing fractal designs (*right*) with lightweight, self-similar elements. Adapted from Rayneau-Kirkhope et al. (2012)

physical and chemical compatibility, and graded structural and compositional variations across the interface. In addition, and as illustrated in Fig. 15, biological composites are represented by fiber hierarchies and geometrical complexities which have inspired more functional models than the traditional fiber bundle model. These have also included fiber branching for pullout suppression inspired by fractal branching (Fig. 12) as illustrated schematically in Fig. 16a, including dendritic and microdendritic phase composites, and the addition of interlayers or transition layers for enhanced compatibility and functionality, including effective wetting at the fiber/matrix interface (Chawla 1996; Kim and Mai 1998; see Fig. 9).

As pointed out by Meyers et al. (2006), the principal defining feature of rigid biological systems in the context of a composite is the existence of mineral and organic components. Table 1 illustrates some of the more prominent systems reviewed in this chapter and in Part VIII: BIOLOGICAL MATERIALS. It can be appreciated that the mechanical properties match a wide range of biological functions which are facilitated by complex, multi-length-scale hierarchies involving multiple interfaces implicit in Fig. 15.

In addition to the pullout suppression provided by branching illustrated schematically in Fig. 16a, pullout caused by interfacial cracking as a result of Poisson ratio effects causing fiber diameter shrinkage can also be presented using fibers having a negative Poisson ratio. Such materials are referred to as *auxetics*, and when stretched, they become thicker normal to the applied stress. Consequently, auxetic reinforcing fibers would thicken with the application of a tensile force. Figure 16b illustrates a simple auxetic hexagon element which becomes square when stretched, compressing material in a direction perpendicular to the stretching force. Figure 16c, d illustrate this effect as it applies to the compression of the fiber/matrix interface for an auxetic fiber/matrix composite.

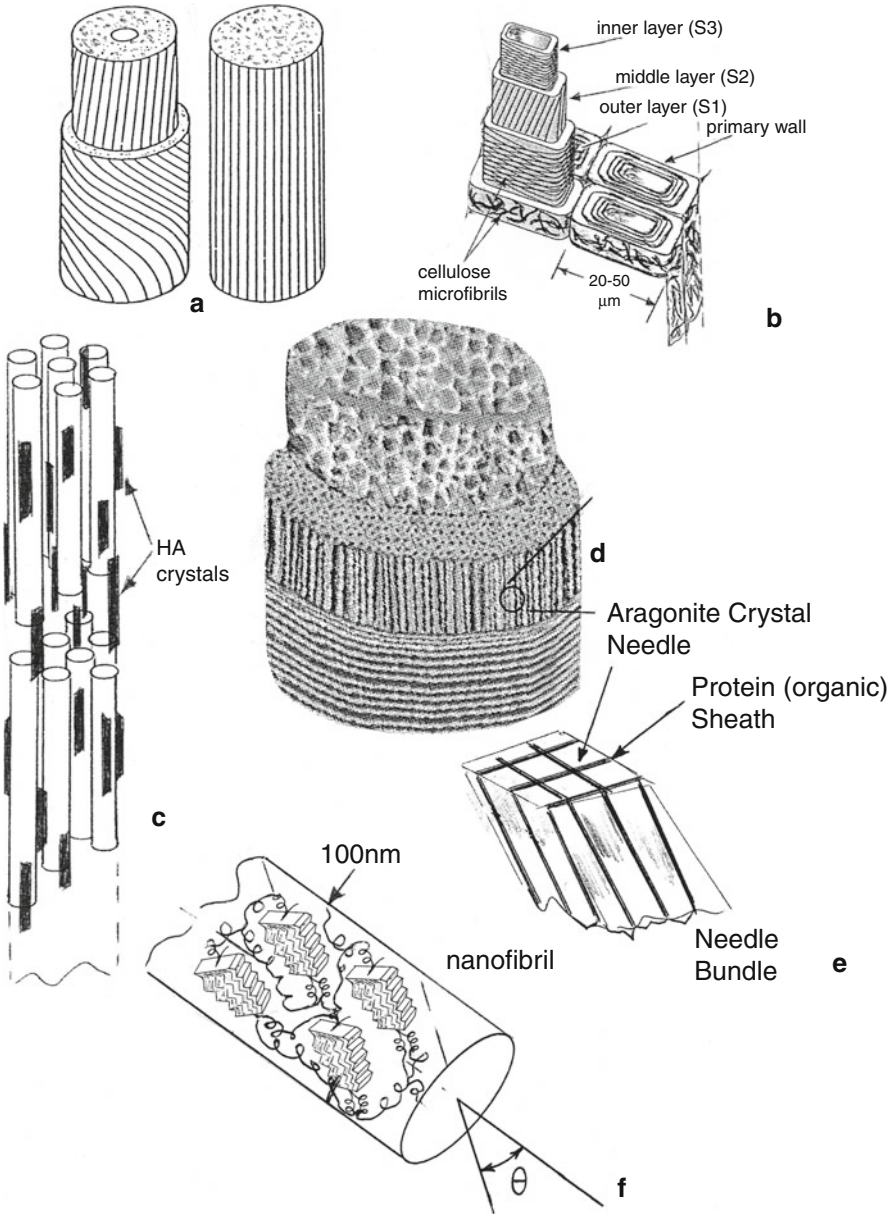


Fig. 15 Comparison (and summary) of biological composite fiber bundle (*left*) and traditional fiber bundle model (*right*) (**a**). (**b**–**f**) show biological fiber composite examples. (**b**) Wood. (**c**) Collagen. HA is hydroxyapatite crystals. (**d**) Feather hierarchy. (**e**) Silkworm silk fiber. (**f**) Conch shell composite

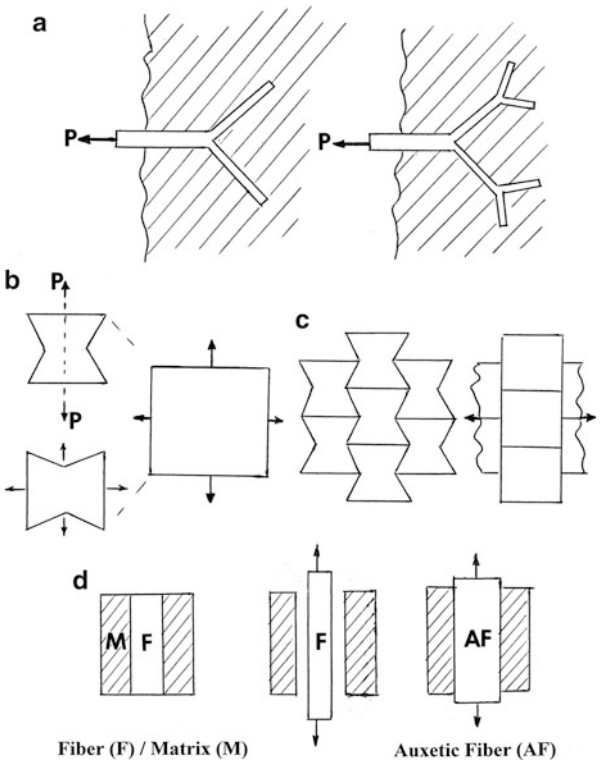


Fig. 16 Fiber pullout for branched fibers (**a**) and auxetic fibers (**b–d**). (**b**) and (**c**) show auxetic element and membrane expansion perpendicular to stretching force (or stress) while (**d**) shows interfacial contraction and expansion comparison

Most conventional materials resist the volume change implicit in Fig. 16b–d by the bulk modulus, K, more than they resist a change in shape, as determined by the shear modulus, G, where

$$G = E/2\,(1 + \upsilon) \tag{8}$$

and from

$$K = E/3\,(1 - 2\upsilon) \tag{9}$$

we can write

$$\left[\frac{(1+\upsilon)}{(1-2\upsilon)}\right] = \frac{3K}{2G} \tag{10}$$

where for conventional engineering materials, the values of K are typically larger than the values of G, which leads to

$$\left[\frac{(1+\upsilon)}{(1-2\upsilon)}\right] \geq \frac{3}{2} \tag{11}$$

Table 1 Mineral and organic components of structural biological composites

Biological composite system	Mineral			Organic				
	Calcium carbonate	Hydroxyapatite	Others	Cellulose	Collagen	Keratin	Chitin	Other protein
Wood				X				
Bones		X			X			
Teeth		X			X			
Shells	X							X
Horns						X		
Beaks						X		
Insect cuticle							X	

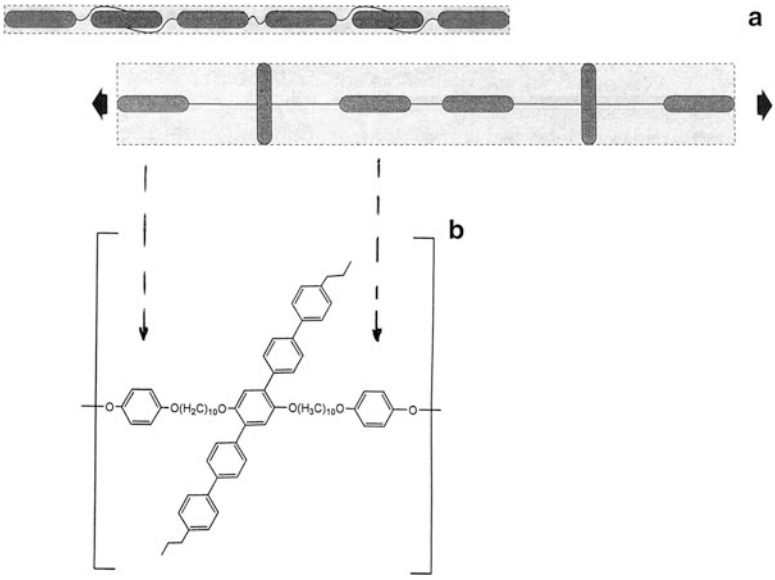


Fig. 17 Examples of molecular auxetics. (**a**) Molecular elements responding to stretching (tensile) force. (**b**) Flexible molecular components

This restricts conventional structural materials to have a Poisson ratio, υ, to be $\upsilon \geq 1/8$. Consequently, $\upsilon \leq 0$ requires the bulk modulus to be much less than the shear modulus ($K << G$). Rewriting Eq. 10 allows one to visualize this phenomenon:

$$\nu = (3K - 2G)/(2G + 6K) \tag{12}$$

There are examples of molecular auxetics. Paper and its compositional cellulose molecules exhibit auxetic behavior since paper of all kinds when stretched in an in-plane direction will expand in the thickness direction. We can illustrate molecular auxetic behavior with reference to Fig. 17 (Yang et al. 2004). Similarly, cellulose molecules (▶ Fig. 11 in chapter "Examples of Natural Composites and Composite Structures") and the nanofibrils composing spider silk (▶ Fig. 17 of chapter "Chemical Forces: Nanoparticles") can be envisioned to exhibit auxetic behavior in some correspondence with Fig. 17. The first example of a synthetic auxetic material was a foam structure (an auxetic foam) (Lakes 1987) although Gibson and Ashby (1982) actually described the auxetic effect in the form of two-dimensional silicone rubber or aluminum honeycomb deforming by flexure of the ribs. Several recent reviews have described developments in modeling, design, and manufacture of auxetic cellular solids and the importance of other auxetic materials in a variety of engineering applications (Alderson and Alderson 2007; Greaves et al. 2011; Prawoto 2012).

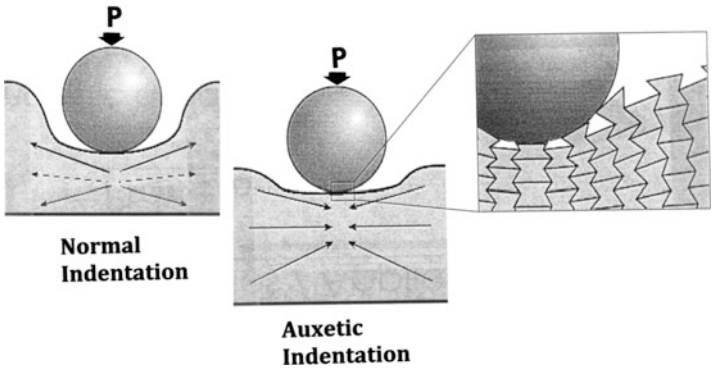


Fig. 18 Auxetic resistance to indentation (Adapted from original by J. N. Grima, University of Malta)

Auxetic materials also include those exhibiting a negative coefficient of thermal expansion (▶ Table 1 in chapter "Examples of Tensor Properties Using Matrix Fundamentals (a physical property)") as well as materials exhibiting resistance to indentation as illustrated schematically in Fig. 18. In normal materials subjected to a traditional hardness test such as Rockwell indentation testing, the hardness expresses a resistance to indentation:

$$H \cong \left[\frac{E}{(1-\upsilon^2)}\right]^m \tag{13}$$

where E is the Young's modulus, υ is the Poisson ratio, and m is a constant equal to 1 for a uniform pressure of indentation. In normal material hardness testing as shown in Fig. 18, the force of the indenter compresses the material which compensates by spreading in directions perpendicular to, and away from, the direction of indentation, indicated by P in Fig. 18. Correspondingly, an auxetic material will compress towards the penetrator (Fig. 18) and this creates a region of denser material which is more resistant to indentation.

It is apparent that auxeticity arises primarily by cooperation between a material's internal geometrical structure or the way this structure deforms when loaded. These structures can derive from "unit cells" as illustrated in the simplest case in Fig. 16b. More complex unit cells can be used to create auxetic foams and other related structures, and these will be described in more detail in Part XI: Advanced Materials Processing and Fabrication: 3D Printing and Additive Manufacturing Technologies. Table 2 lists a range of non-auxetic (normal) and auxetic materials and their corresponding Poisson ratio. Some of these non-auxetic materials are derived from ▶ Table 2 in chapter "Examples of Tensor Properties Using Matrix Fundamentals (a physical property)". It can also be noted in Table 2 that there are a number of natural auxetic materials such as zeolites (such as thomsonite: $Na_4Ca_8[Al_{20}Si_{20}O_{80}]\cdot 24H_2P$ (Grima et al. 2009)) and pyrolytic graphite which is directionally auxetic along the c-axis. It has also been observed that many cubic

Table 2 Poisson ratio for non-auxetic and auxetic materials

Non-auxetic materials	Poisson ratio (υ)
Rubber	0.50
Ag, Br	0.40
Pb, Nb	0.39
Al	0.35
Glass	0.30
Fe (and steels), Al_2O_3	0.27–0.29
MgO, SiC	0.14–0.17
Cellular solids (polymer foams)	0.1–0.4
Be	0.1
Diamond	0.07
Cork	0.0
Auxetic Materials	Poisson ratio (υ)
Zeolites	−0.33 to −0.55
Auxetic polyurethane and other polyester foams	−0.3 to −0.7
Pyrolytic graphite (c-axis)	−0.21
SiO_2 (α – cristobalite)	−0.2 to −0.5

metals when stretched along the [110] direction exhibit a negative Poisson ratio. We will explore auxetic materials in more detail as they are applied in novel biomaterials and biomedical innovations later in chapter "▶ Strategies for Bone Replacement and Tissue Augmentation."

Protein Factories and Related Concepts

Perhaps the ultimate goal in materials science and engineering involves the creation of materials or materials systems with specific properties, behavior, performance, or functionality based on inclusive fundamentals and atomic and molecular structures. In the case of natural or biological materials, this involves the genome or the entirety of an organism's hereditary information which is encoded in the DNA or RNA. The genome includes both the genes and the noncoding sequences of the DNA/RNA.

Synthetic biology creates biological materials and materials systems by mimicking biological functionality through DNA strand building rather than altering existent DNA strands as is the case for genetic engineering. This approach combines biological science fundamentals with engineering as biological materials science and engineering. The new DNA strand is then placed in an empty living cell (a bacterium or virus) which can be built to some biological materials specification. As we illustrated earlier in this chapter, virus cells and bacteria can be manipulated to produce useful protein materials. This approach creates protein factories or cell factories which can manufacture a wide range of useful biological products, including drugs and related therapies. Single cells can also be encrypted in various ways or the components within a cell can be selectively programmed as part of a process in gathering or processing information or responding to some form

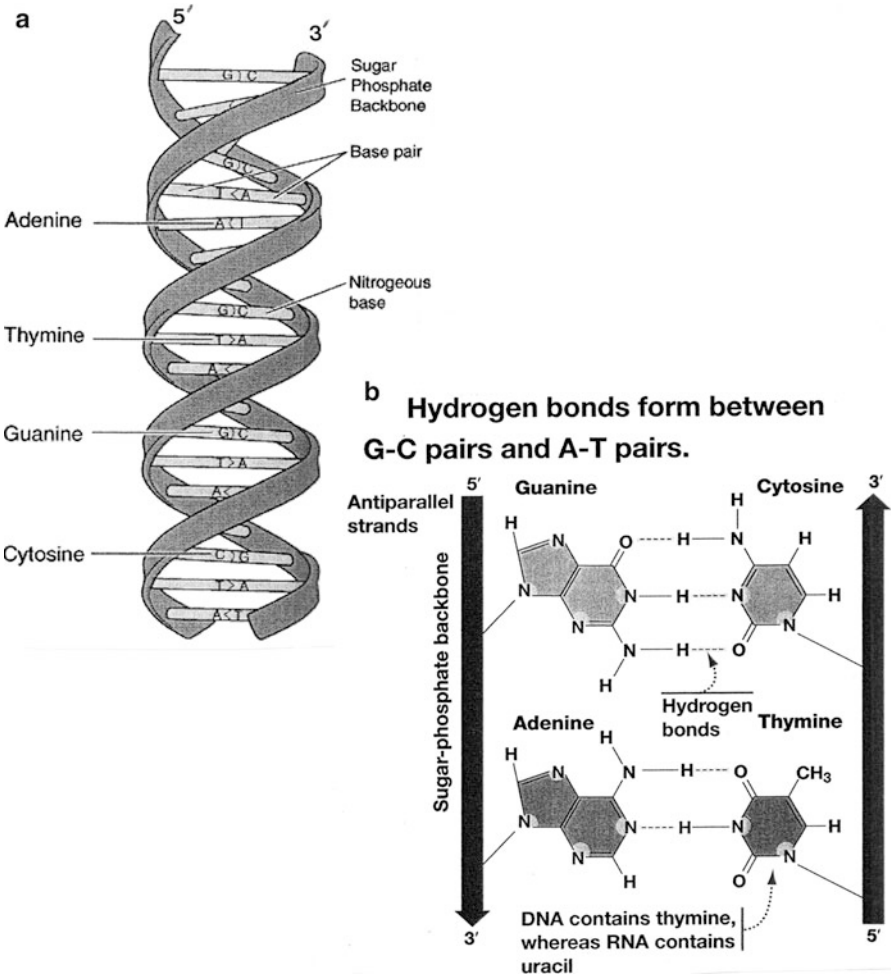


Fig. 19 DNA strand structure (**a**) and composition schematics (**b**) (Courtesy of Pearson Education, Inc.)

of incoming signals. This concept can be adapted to creating synthetic biology devices as a form of distributed computation (Macia et al. 2012).

Figure 19 illustrates the two strands of DNA: one that is copied into mRNA (messengerRNA) and subsequently translated to have the complementary sequence to the mRNA and the other base sequence of the opposite strand which directly corresponds to the codons in the mRNA. The template strand refers to the DNA sequence that is copied during mRNA synthesis. The opposite strand is called the coding strand or the mRNA-like strand because the sequence corresponds to the codons that are translated into protein. There are 40 million positions in the human genome with DNA sequences that are different from those in primates. The human genome has a total of three billion pairs.

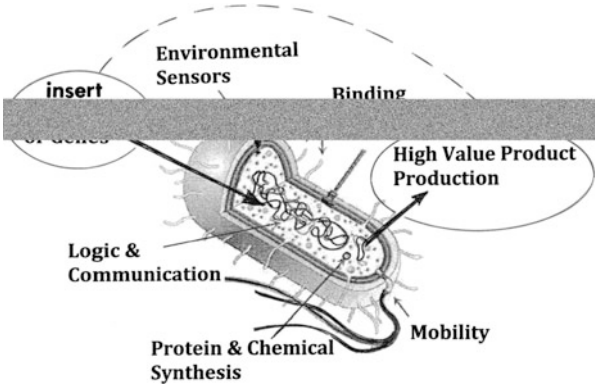


Fig. 20 Cellular factories in synthetic biology. DNA strands or genes are introduced to program cells to become cellular factories in order to produce high-value products. Cellular functions are indicated generally

It can be noted in Fig. 19 that each strand is composed of a row of bases (adenine (A), thymidine (T), cytosine (C), and guanine (G)) bound to a sugar ring (deoxyribose) and to a phosphate PO_4^{-2} group: base + sugar + phosphate group forming a nucleotide. The two strands run in opposite directions and are antiparallel: one backbone being 3′ (three prime) and the other 5′ (five prime), which refers to the direction of the third and fifth carbon on the sugar molecule. Within human cells, DNA is organized into long structures called chromosomes which are duplicated (by DNA replication) during cell division. The largest human chromosome (chromosome number 1) is ~220 million base pairs long.

A gene is a sequence of DNA that contains genetic information, and within a gene, the sequence of bases along a DNA strand (Fig. 19) defines an mRNA sequence which then defines one or more protein sequences. The relationship between nucleotide sequences of genes and the amino acid sequences of proteins is determined by the so-called rules of translation or the genetic code which consists of *codons* formed from a sequence of three nucleotides, e.g., ACT, CAG, TTT, etc. These codons are copied into mRNA by RNA polymerase which is then decoded by a ribosome that reads the RNA sequence. Since there are four bases (A, T, C, G) in three-letter combinations as noted above, there are 4^3 (or 64) combinations. These encode the 20 standard amino acids shown in ▶ Fig. 1 in chapter "Structure and Function of Viruses and Bacteria", giving most amino acids more than one possible codon. The end of coding regions is denoted by codons TAA, TGA, and TAG.

As we noted previously in ▶ Fig. 5 (chapter "Structure and Function of Viruses and Bacteria"), *Escherichia coli* can serve as a protein factory where by attaching sugars to proteins large quantities of different glycoproteins can be generated for various medical and biological studies. In addition, and as illustrated in Fig. 20, many other engineered components and systems can be achieved by identifying the various organism functions and manipulating them in a fundamental biochemical process or processes, often referred to as engineered bacteria. Indeed, synthetic biology is often referred to as the design and construction of biological devices often by inserting man-made DNA into cells or organisms (especially bacteria). The "Registry of Standard Biological Parts" (Catalog-partsregistry.org;

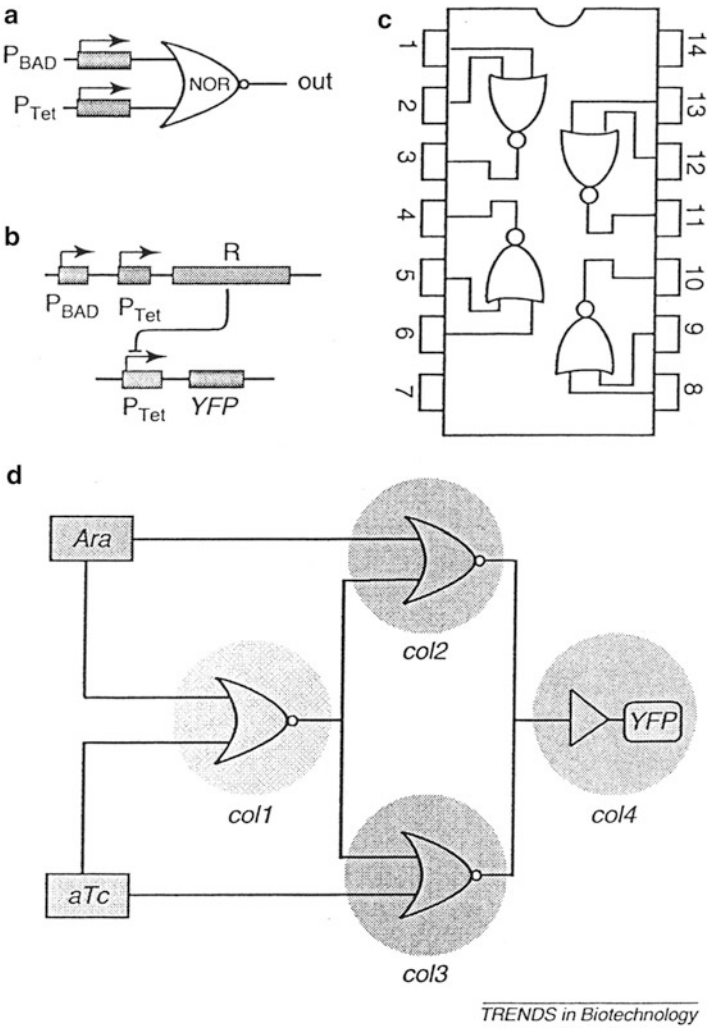


Fig. 21 Spatially segregated colonies of engineered strains connected through quorum wiring molecules (acyl homoserine lactones). Two orthogonal NOR gates are combined to build all two-input one-output logic gates. (**a**) Basic units constructed on *E. coli* using promoters ρ_{BAD} and ρ_{Tet} which are activated under the presence of arabinose (*Ara*) and anhydrotetracycline (*aTc*). Repressor transcription is shown in (**b**). (**c**) shows a spatially extended combination of NOR gates. (**d**) Corresponding molecular logic gates (two-input one-output logic gates) (Adapted from (Macia et al. 2012)

http://parts.mit.edu) includes specially designed DNA sequences for use in two kinds of bacteria, yeast (which is more complex than bacteria), bacteriophage T7, and mammalian cells. J. C. Venter and colleagues (Gibson et al. 2010) have described the design and synthesis of 1.08 million base pairs which were transplanted into *Mycoplasma capricolum* recipient cell to make new *Mycoplasma mycoides* cells controlled only by the synthetic chromosome.

Such manipulations of cells have led to gene synthesis, vaccine and antibody production, industrial enzymes, and prospects for creating biofuels using a DNA sequence inserted into a host organism that would carry out the biofuel production process. Pharmacologically important molecules like insulin have in fact been manufactured by recombinant DNA technology in bacterial cell engineering for more than three decades (Goeddel et al. 1979). With the ability to create artificial genomes (Gibson et al. 2010), prospects for customized drug-producing or therapeutic bacteria become viable biomaterial avenues.

Finally, and in retrospect, the prospects for fabricating complex synthetic circuits as reviewed by Macia et al. (2012) and illustrated schematically in Fig. 21 deserve some comment. Such "living circuits" can provide a distributed computing platform for achieving complex computations and prospects for easy reprogramming by replacing a few cell types by others or modifying the specific logic function, making it possible to construct a wide range of diverse circuits as illustrated in Fig. 21.

References

Alam R, Fontaine DM, Branchini BR, Maye MM (2012) Designing quantum rods for optimizing energy transfer with firefly luciferase enzymes. Nanoletters 12(6):3251–3256

Alderson A, Alderson K (2007) Auxetic materials. J Aerosp Eng 221:565–575

Armstrong R, Spiller N (2011) Synthetic biology: living quarters. Nature 607:916–918

Arzt E, Gorb S, Spolenck R (2003) Evidence for van der Waals adhesion in gecko setae. Proc Natl Acad Sci U S A 100:10603–10606

Autumn K, Grovich N (2008) Gecko adhesion evolutionary nanotechnology. Philos Trans R Soc 366:1575–1590

Bar-Cohen Y (ed) (2006) Biomimetics: biologically inspired technologies. Taylor & Francis (CRC Press), Boca Raton

Baumgartner W, Fidler F, Weth A, Habbecke M, Jacob P, Butenweg C, Bohme W (2013) Investigating the locomotion of the sandfish in desert sand using NMR-imaging. Plos One 3(10):e 3309

Bay A, Cloetens P, Suhonen H, Polvigneron J (2013) Improved light extraction in the bioluminescent lantern of a *Photuris* firefly (Lampyridae). Optics Express 21(1):764

Benyus JM (1997) Biomimicry: innovation inspired by nature. Morrow Publishers, New York

Berthe RA, Westhoff G, Blackmann H, Gorb SN (2009) Surface structure and frictional properties of the skin of the Amazon tree boa *Corallus hortulanus* (Squamata, Boidae). J Comp Physiol A 195(3):311–318

Bruet BJF, Song J, Boyle MC, Ortiz C (2008) Materials design principles of ancient fish armour. Nat Mater 7:748–756

Burns JN, Lenaghan SC, Zhang M, Stewart CN (2012) Nanoparticle biofabrication using English ivy (*Hedra helix*). J Nanobiotechnol 10:41–50

Chawla KK (1996) Interfaces in metal matrix composites. Compos Interfaces 4(5):287–298

Contagal CH, Bachmann MH (2002) Advances in *in-vivo* bioluminescence imaging of gene expression. Ann Rev Biomed Eng 4:235–260

Gather MC, Yun S-H (2011) Cellular lasers. Nat Photon 5:438–441

Geim AK, Dubonas SV, Grigoresa IV, Novoselov KS, Zhokos AA, Shapoval SY (2003) Microfabricated adhesive mimicking gecko foot-hair. Nat Mater 2:461–463

Gibson LJ, Ashby MF (1982) The mechanics of three-dimensional cellular materials. Proc R Soc Lond A Math Phys Sci 383:43–59

Gibson DG, Glass JI, Latigue C, Naskov VN, Chuang RY, Agire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishna Kumar R, Asasad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi Z-O, Segall-Shapiro TH, Calvey CH, Pamar PP, Hutchison CA, Smith HO, Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science (published online 20 May 2010).(doi:10.1126/science.1190719

Goeddel DV, Kleid DG, Boliver F, Heyneker HC, Yansura DG, Crea R, Hirose T, Kraszewski A, Itakura K, Riggs AD (1979) Expression in *Escherichia coli* of chemically synthesized genes for human insulin. Proc Natl Acad Sci U S A 76:106–110

Greaves G, Greer A, Lakes R, Rouxel T (2011) Poisson's ratio and modern materials. Nat Mater 10:823–837

Grima JN, Cassar RN, Gatt R (2009) On the effect of hydrostatic pressure on the auxetic character of NAT-type silicates. J Non-Cryst Solids 355:1307–1312

Havlin S, Buldyrev SV, Goldberger AL, Mantegna RN, Ossadnik SM, Peng C-K, Simons M, Stanley HE (1995) Fractals in biology and medicine. Chaos Solut Fract 6:171–201

Hepburn HR, Ball A (1973) On the structure and mechanical properties of beetle shells. J Mater Sci 8:618–623

Hergenrother WL, Shultz LL, Lin CJ (2012) Keratin in rubber applications. US patent application number 20120329916

Karthaus O (2013) Biomimetics in photonics. CRC Press (Taylor & Francis Group), Boca Raton

Kim J-K, Mai Y-W (1998) Engineered interfaces in fiber reinforced composites. Elsevier Sciences, New York

Kim J-J, Lee Y, Kim HG, Choi K-J, Kweon H-S, Park S, Jeong K-H (2012) Biologically inspired LED lens from cuticular nanostructures of firefly lantern. Proc Natl Acad Sci USA 109(46):18674–18678

Kinoshita S, Yoshioka S, Miyazaki J (2008) Physics of structural colors. Rep Prog Phys 71:1–30

Klonowski W, Stepien R, Stepien P (2010) Simple fractal method of assessment of histological images for application in medical diagnostics. Nonlinear Biomed Phys 4:7 (8 p)

Lakes R (1987) Foam structures with a negative Poisson's ratio. Science 235:1038–1040

Large M (ed) (2013) Optical biomimetics: materials and applications. Woodhead Publishing, Cambridge, UK

Lee H, Lee BP, Messersmith PB (2007) A reversible wet/dry adhesive inspired by mussels and geckos. Nature 448:338–341

Lewis SM, Cratsley CK (2008) Flash signal evolution, male choice, and predation in fireflies. Ann Rev Entomol 53:293–321

Lin AYM, Brunner R, Chen P-Y, Talke FE, Meyers MA (2009) Underwater adhesion of abalone: the role of van der Waals and capillary forces. Acta Mater I57:4178–4185

Macia J, Posas F, Sole RV (2012) Distributed computation: the new wave of synthetic biology devices. Trends Biotechnol 30(6):8 p

Mandelbrot BB (1982) The fractal geometry of nature. W.H. Freeman, New York

Mandelbrot BB (2012) The fractalist: memoir of a scientific maverick. Pantheon Books/Random House, New York

Martines E, Seunarine K, Morgan H, Gadegaard N, Wilkinson CDW, Richle O (2005) Superhydrophobicity and superhydrophilicity of regular nanopatterns. Nano Lett 5:2097–2103

McNamara ME, Briggs DEG, Orr PJ, Nuh H, Cao H (2012) The original colors of fossil beetles. Proc Roy Soc B: Biol Sci 279(1731): 1114–1121

Meizer B, Steinbrecher T, Seidel R, Kraft O, Schwaiger R, Speck T (2010) The attachment strategy of English ivy: a complex mechanism acting on several hierarchical levels. J R Soc Interfaces 7:1387–1389

Meyers MA, Lin AYM, Seki Y, Chen P-Y, Kad BK, Bodde S (2006) Structural biological composites: an overview. JOM 58(7):35–41

Parker AR, Lawrence CR (2001) Water capture by a desert beetle. Nature 414:33–34

Prawoto Y (2012) Seeing auxetic materials from the mechanics point of view: a structural review on the negative Poisson's ratio. Comput Mater Sci 58:140–153

Prum RO, Torres R, Kokach C, Williamson S, Goodman SM (1999) Coherent light scattering by nanostructured collagen arrays in the caruncles of the malagasy asities (Eurylaimidae: ares). J Exp Biol 202:3507–3522

Ralston E, Swain G (2011) Can biomimicry and bioinspiration provide solutions for fouling control? Mar Technol Soc J 45:216–227

Rayneau-Kirkhope D, Mao Y, Farr R (2012) Ultralight fractal structures from hollow tubes. Phys Rev Lett 109:204301

Schmitz H, Murtz M, Blackmann H (1997) Infrared detection in a beetle. Nature 386(6627): 773–774

Seago AE, Brady P, Vigneron J-P, Shultz TD (2009) Gold bugs and beyond: a review of iridescence and structural color mechanisms in beetles (Coleoptera). J R Soc Interface 6: S165–S184

Shimomura M (2010) The new trends in next generation biomimetics material technology: learning from biodiversity. Quart Rev 37:53–75

Tsujii K (2009) Superhydrophobicity and superhydrophilicity: mechanism and application. Yuneda Publishing, Tokyo

Watanabe K, Hoshino T, Kanda K, Haruyama Y, Matsai S (2005) Fabrication of Morpho-butterfly-scale quasistructure by focused ion beam chemical vapor deposition. J Appl Phys 44:L48–L60

Wilson SJ, Hutley MC (1982) The optical properties of moth eye antireflection surfaces. J Mod Opt 29(7):993–1009

Yang W, Zhong-Ming L, Shi W, Xie B, Yang W-B (2004) Review on auxetic materials. J Mater Sci 39:3269–3279

Youngquist JA, English BE, Scharmer RC, Chow P, Shook SR (1994) Literature review on use of nonwood plant fibers for building materials and panels. U.S. Department of Agriculture, Forest Products Laboratory general technical report FPL-GTR-80

Yu J, Chary S, Das S, Tameller J, Pevika NS, Turner KL, Israelachvilli JN (2001) Gecko-inspired dry adhesive for robotic applications. Adv Funct Mater 21(16):3010–3018

Yurdumankan B, Raravikar NR, Ajayan PM, Dhinojwala A (2005) Synthetic gecko foot-hairs from multiwall carbon nanotubes. Chem Comm 5(30): 3799-3801

Zhao J, Wang A, Campbell P, Green MA (1999) 19.8% efficient honeycomb multicrystalline silicon solar cell with improved light trapping. IEEE Trans Electron Devices 46(10): 1978–1983

Part X

Biomaterials

Strategies for Bone Replacement and Tissue Augmentation

Contents

Abstract

Following a brief description and review of hydroxyapatite-reinforced polymeric and related systems for bone replacement strategies, the use of biodegradable metals and polymers is discussed as temporary bone support systems to allow bone regrowth and healing, including screws. The use of these materials for a variety of nonpermanent stents is described in the context of novel stent structures, including shape memory materials and the fabrication of auxetic structures to achieve stent shapes.

Introduction

This chapter begins with an overview of efforts to create requisite material regimes to facilitate bone tissue repair and replacement in circumstances where repair is intractable. As discussed in Part VII, chapter "► Eutectic Composites", the bone is a composite of collagen (protein) strands and hydroxyapatite which provides strength as a consequence of its high elastic modulus but still allows for flexure of the collagen strands.

In cases where bone fractures require stabilization to allow healing by regrowth of bone tissue, the placement of strong, rigid structures and their fastening using screws can be facilitated by using biodegradable materials which are reacted and dissolved in the body after the bone has healed; magnesium plates, rods, and screws

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_31

along with suitable alloys of magnesium have been clinically utilized for decades. These materials are augmented by a variety of biodegradable polymeric materials which, because of their lower strength, are not as reliable or applicable as lightweight metals and alloys such as magnesium and its alloys. Nonetheless, biodegradable polymeric materials along with magnesium and its alloys are utilized effectively in stent structures which are intended to be temporary, having some medical utility only until healing of attached tissue occurs.

In the case of permanent stent materials, shape memory alloys are useful in affecting some appropriate and extended functions, and other biomedical functions are also achieved using shape memory alloys, including dental braces and the like. While heat can affect shape memory biomedical devices, the use of auxetic structures in stents or stent design can allow shape features to be achieved through simple stress application such as pulling of the stent after it has been placed in some tissue conduit.

These examples represent a few novel applications of materials and materials structures in biological and biomedical venues, especially those involving human health and quality of life issues.

As noted at the outset of Part VIII, biomaterials include a wide range of materials and materials systems, including composites as well as materials innovations which mimic biological systems and which are applied in biomedical and related situations. This chapter begins with bone replacements or substitutes which include biodegradable systems, although the area of biodegradable biomedical materials applications includes a separate and broad area involving biomaterials.

Bionics of Bone and Bone Structures

As noted in chapter "▶ Examples of Natural Composites and Composite Structures" (Fig. 1), the bone is a viscoelastic composite capable of self-repair. Its mechanical properties in particular depend upon cortical porosity, the size and distribution of hydroxyapatite (HA) crystals within collagen fibrils, and the presence and distribution of microcracks. The crystalline HA structure provides compressive strength and brittleness, while the collagen fibril matrix provides toughness and tensile strength.

Within the last decade, bone tissue repair accounted for roughly 0.5 million surgical procedures per year in the United States (Geiger et al. 2003). While natural repair mechanisms involving angiogenesis, osteogenesis, and chronic wound healing occur in the body, there are circumstances where tissue regeneration cannot accommodate the healing process and require clinical intervention using biomaterials or bone substitutes which allow either a temporary porous scaffold or a permanent device that reduces the defect size and promotes healing through the creation of new tissue. In this regard, optimal bone substitutes involve a device or regime having similar mechanical and biological properties, i.e., a collagen–HA composite. Collagen can be extracted from a variety of tissue regimes, including animals, while HA can be extracted from bones, especially cortical bovine bone, or obtained commercially with different, controlled crystal sizes (Wahl and Czernuszka 2006).

Xu et al. (2012) have developed a technology for producing bone composite from fish scales and HA, the fish scales providing the collagen matrix. Methods employed to fabricate bone substitutes include a so-called sponge replica method which produces a highly porous interconnected structure that mimics cancellous bone or electrospinning to produce fibrous scaffolds or combination (Kim and Lee 2011).

In addition to HA ($(Ca_{10}\ CaPO_4)_6\ (OH)_2$), tricalcium phosphate (TCP: $Ca_3(PO_4)_2$) and biphasic calcium phosphate, BCP, a mixture of HA and TCP, are common bone replacement materials which are often reinforced by zirconia (ZrO_2). Polymer-based biomaterials are also widely used. These include polymethyl methacrylate (PMMA) and polycaprolactone (PCL): PMMA–PLC–HA or ZrO_2/BCP scaffolds wrapped with HA-loaded PMMA–PCL fibers, etc. W. Bonfield and colleagues also pioneered the use of HA-reinforced high-density polyethylene (HA/HDPE) composites in the 1980s but were found to be unsuitable for load-bearing requirements (Bonfield et al. 1981; Wang and Bonfield 2001). More recently, Shen et al. (2011) have demonstrated the potential for HA/Vectra A950 (HA/VA) composite prepared by twin-screw extrusion and injection molding as a bone replacement biomaterial because the VA (a copolymer of *p*-hydroxybenzoic acid and 6-hydroxy-2-naphthoic acid commercial thermotropic liquid crystal polymer) has a tensile modulus, strength, and strain at break of 10.6 GPa, 0.18 GPa, and 3.4 %, respectively. This is in contrast to cortical bone with a modulus between ~10 and 30 GPa, a strength of 0.05–0.15 GPa, and a strain to break of roughly 1–3 %. Polyacrylic acid (PAA) is also added to this composite to improve interfacial adhesion.

The creation of specific bone replacements and promising scaffold configurations has recently benefited through the use of 3D printing or solid freeform fabrication (SFF) techniques which are described in detail in Part XI (Hufmacher et al. 2004). These techniques have the ability to fabricate patient-specific prototypes using CT scans and CAD (computer-aided design) modeling, which allows even collagen–HA scaffolds to be produced by controlling variables at several length scales. Unfortunately, the structure of SFF-fabricated biocomposite prototypes is significantly different from the hierarchical structures of the natural bone which is a nanosized HA and highly oriented collagen fibers at the nanoscale (as implicit in ▶ Fig. 2 in chapter "Examples of Natural Composites and Composite Structures"). In this context, Shen et al. (2011) have recently concluded that "Therefore, there is not much progress in developing polymer matrix composites for load bearing bone replacements, and metals may still be the best at present." Metal implants will be discussed later in chapter "▶ Implant Materials and Structures."

Biodegradable Biomedical Materials and Applications

Biodegradable (biomedical) orthopedic implant applications during the healing process in the human body have involved sutures and suture anchors, screws, staples, pins, and rods to support weakened or fractured areas, especially involving bone injuries or osteoporosis. Magnesium, magnesium alloys, and polymeric

materials (both natural and synthetic) form the more important biodegradable materials. Magnesium and magnesium alloys provide higher mechanical strength and stiffness which may be especially important in orthopedic procedures where high loads are involved. On the other hand, degradable polymers are extremely effective for drug encapsulation for controlled release in the body. One of the first recorded uses of a biodegradable material involves the use of magnesium wires as ligatures to stop vessel bleeding as described in an 1878 paper by Edward C. Huse (1878). Huse in fact observed that complete degradation (or dissolution) of the Mg wire depended on the wire diameter.

Sutures also were among the earliest applications of biodegradable polymer fibers such as catgut (natural fiber in animal intestine walls) used by the Romans, but at the expense of inflammatory response. Even earlier surgical suture evidence using flax, hemp, and cotton fibers extends to circa 3,000 BC in Egypt. Correspondingly, the first biodegradable synthetic polymer suture, poly(glycolic acid), was approved in 1969, while poly(lactic acid), an improved suture polymer, was approved in 1972. Polydioxanone was developed in the 1980s for biodegradable sutures and bone pins, while polyanhydride implantable drug delivery systems were developed around 1996.

Biodegradable materials have the advantage of not having to be removed upon achieving their purpose, degradation products are metabolized or excreted, and in the case of polymer encapsulations, effective drug release strategies can be achieved. The removal of nondegradable devices or appliances increases the cost and risk of morbidity. In addition, long-term implants can cause complications, including infection or mechanical failure which would normally require corrective surgery. In tissue engineering, biodegradable polymers can be designed to approximate tissue, providing a polymer scaffold that can withstand mechanical stresses and provide suitable cell attachment surface features for growth and degrade at a rate that will allow the load to be transferred to the new tissue. In this respect, biodegradable polymer devices also eliminate stress shielding issues involving the bone because their Young's modulus is often compatible with the bone.

Figure 1 illustrates the biodegradable polymer stents and screws and their application. Stents are often used to allow damaged tissue vessels to heal and are inserted through a natural or surgical orifice. The degradation of the device negates having to retrieve it. Biodegradable polymer screws in healing bone (Fig. 1b) exhibit the same advantage and allow the bone to maintain a more natural function.

Among the biodegradable polymers, aliphatic polyesters [poly(hydroxyl acid)s] such as poly(lactic acid) or polylactide (PLA), poly(glycolic acid) or polyglycolide (PGA), and poly(ε-caprolactone) (PCL) and their copolymers have been prominently used in implantable devices (Fig. 1). Figure 2 shows the structures of their monomers (Ohya et al. 2011).

Polymers degrade primarily by hydrolysis and oxidation mechanisms. Hydrolysis is a second-order reaction that depends on the concentration of vulnerable bonds and water, which are partially soluble in many polymers. Oxygen destroys polymer chains and can also react in the form of other oxidative species such as peroxides or

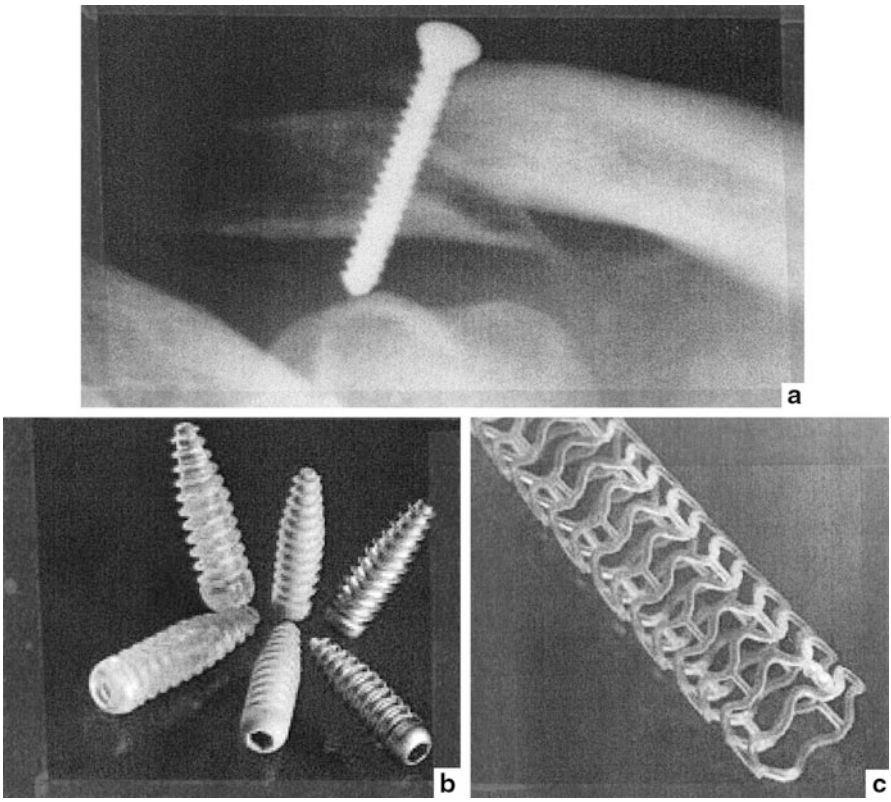


Fig. 1 Examples of biodegradable polymer biomedical devices and applications. (**a**) Degradable polymer screw bond fixation. (**b**) Inferential screws made of poly(lactic acid) polymer matrix with hydroxyapatite addition. (**c**) Arterial drug-eluting stent consisting of biodegradable PLLA (poly-L-lactic acid) backbone with poly-D,L-lactide bioabsorbable polymer coating that controls the release of antiproliferative drug, everolimus (Courtesy of Abbott Laboratories)

free radicals. Natural enzymes produced by the body can also catalyze these reactions. In addition, inflammatory cells migrating to the implant site can produce peroxides. In some cases, hydrolysis and oxidation can act simultaneously and may in fact reinforce one another.

As described in a brief history of biodegradable magnesium implants by Frank Witte (2010), the use of magnesium for biomedical applications began shortly after its discovery by Sir Humphry Davy in 1808. As noted earlier, ligature wires were among the first biomedical applications followed by tubes for vessel connections and rods, plates, and screws for bone fixtures. During the last decade, cardiovascular magnesium alloy stents similar to those shown conceptually in Fig. 1 have been utilized where temporary stents are required in assisting arterial remodeling and healing in contrast to permanent metallic alloy stents which are normally fabricated from type 316 stainless steel (Morevej and Mantovani 2011).

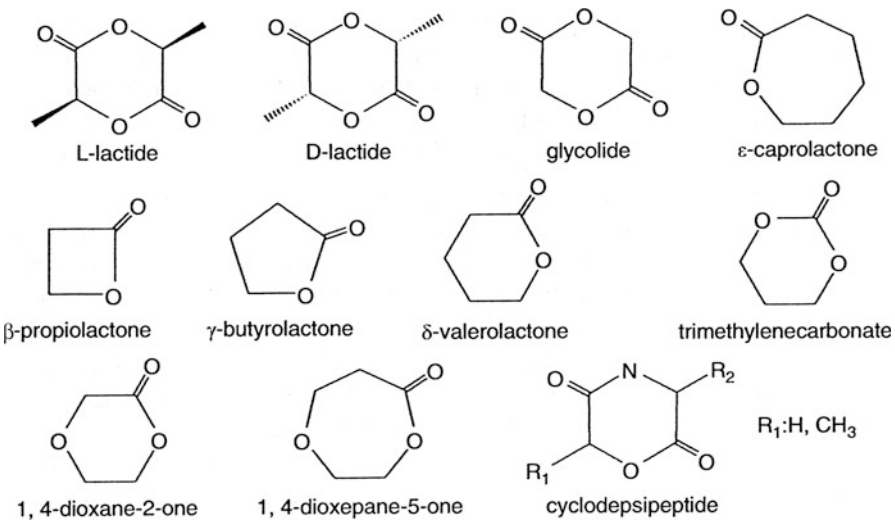


Fig. 2 Cyclic monomers of biodegradable aliphatic polyester polymers

The first biodegradable metallic stent was fabricated from Armco iron and implanted in the rabbit aorta in 2001 (Peuster et al. 2001), while most contemporary magnesium alloy stents are coated with a biodegradable polymer such as PLA to control the metallic biodegradation rate.

Generally, magnesium and its alloys have low density (around 2 g/cm^3, which is ~1/3 as dense as Ti alloys and 1/2 as dense as stainless steel or Co–Cr alloys) and elastic modulus nearly compatible with the cortical bone: E ≅ 12 GPa for pure Mg to 38 GPa for Mg alloys. Consequently, the use of biodegradable Mg and Mg alloys (containing rare earth elements and other metallic elements including Li, Zr, Ca, Zn, and Al) in biodegradable bone implant applications is a major biomedical materials area (Fig. 3). Figure 3 shows a complex, biodegradable Mg implant composed of screws and plate. Lithium in Mg enhances ductility, while Zr serves as an effective grain-refining agent. Lithium additions also change the crystal structure from Mg-hcp to bcc. Table 1 illustrates some principal biodegradable Mg alloy compositions and their major applications.

Magnesium alloys are degraded by corrosion in aqueous regimes by several oxidation–reduction reactions which are influenced by the alloying elements. The common half-cell reaction:

$$\mathrm{Mg\ (solid)} + 2\mathrm{H_2O} \rightarrow \mathrm{Mg\ (OH)_2(solid)} + \mathrm{H_2(gas)} \tag{1}$$

Contributing half-cell reactions include:

$$\mathrm{Mg\ (solid)} \rightarrow \mathrm{Mg^{2+}(aqueous)} + 2\mathrm{e\ (oxidation-reduction)} \tag{2}$$

Fig. 3 Example of complex biodegradable magnesium alloy (Mg–Ca) implant (X-ray). A second operation to remove the implant system would not be required after bone regeneration and implant dissolution (Courtesy of Helmholtz Zentrum Geesthacht, Germany)

Table 1 Common biodegradable magnesium alloys and their biomedical applications

Alloy[a]	Alloying elements (wt. %)	Applications
Mg	–	Ligature wire, vessel connector tubes, screws, plates, rods
AZ 31	3Al, 1Zn	Orthopedic implants (biodegradable)
AZ 63	6AL, 3Zn	Biodegradable stents
AZ 91	9Al, 1Zn	Screws, plates
Mg. x Ca	x = 1, 2, 3 (Ca)	Implant plates and rods
Mg–Zn–Ca	1Zn, 1Ca	Implant plates and rods
Mg–Zn–M_n	1Zn, 1M_n	Bone implant plates and rods
LAE 442	4Li, 4Al, 2RE[b]	Degradable stent
WE 43	4Y, 3RE[b]	Orthopedic implants
ZE 41	7Zn, 1RE[b]	Orthopedic implants, stents
AE 44	4Al, 4RE[b]	Plates, screws/implant
WZ 21	2Y, 1Zn	Slow-degrading implants, pins, and screws
Mg x Gd	x = 5, 10, 15 (Gd)	Biodegradable stent

[a]Based on data in Razavi et al. (2012)
[b]*RE* rare earth element: Nd, Ho, Er, Lu, Tb, Tm

$$2H_2O + 2e \rightarrow H_2(\text{gas}) + 2OH\ (\text{aqueous : reduction reaction}) \quad (3)$$

$$Mg^{2+}(\text{aqueous}) + 2OH \rightarrow Mg\ (OH)_2(\text{solid}) \quad (4)$$

or

$$Mg^{+} + H_2O \rightarrow Mg^{+} + OH + H_2(\text{gas}) \quad (5)$$

With Zn as an alloying agent, the following additional reactions would occur:

$$Zn + 2H_2O \rightarrow Mg\ (OH)_2 + H_2 \uparrow \quad (6)$$

$$Zn + Zn^{2+}(\text{aqueous}) + 2e\ (\text{oxidation reaction}) \quad (7)$$

$$Mg + Zn^{2+}(\text{aqueous}) \rightarrow Zn + Mg^{2+}(\text{aqueous}) \quad (8)$$

It can be noted that Mg reactions produce hydrogen gas, but Zn additions tend to reduce the hydrogen gas emission significantly. While in some applications and in circumstances where hydrogen gas emission may be elevated due to increased rate of biodegradation (corrosion), some subcutaneous gas cavities can form, although in these cases most patients have experienced no pain or infections. In the case of arterial stents, hydrogen gas evolution is carried away in blood flow, and as noted, biodegradable polymer coatings can control the hydrogen emission by reducing the rate of corrosion of the Mg alloy stent. Hornberger et al. (2012) have also recently discussed other biomedical coatings on magnesium alloys such as calcium phosphate.

Magnesium and magnesium alloys are not toxic since tissue is often rich in magnesium and relatively large Mg concentrations are tolerated by the body.

Biomedical Applications of Shape Memory Materials

Materials able to recover their original shape after being severely and quasi-plastically deformed or distorted when subjected to an applied stimulus, such as heat or light, are referred to as shape memory materials, or smart materials, since they in effect "remember" their original shape. This phenomenon is known as the *shape memory effect* and has been observed in a number of materials systems, including alloys, polymers, and ceramics (Huang et al. 2010; Otsuka and Wayman 1998).

Polymers with shape memory effect properties are called *shape memory polymers*. Common polymers such as polymethyl methacrylate (PMMA) exhibit a heat-responsive shape memory effect as illustrated in Fig. 4. As shown in Fig. 4a, a rectangular piece of PMMA is heated above its glass transition temperature (T_g) and deformed into the shape shown in ▶ Fig. 60b of chapter "Chemical Forces: Nanoparticles," after which it cools below T_g. Heating the cooled shape in Fig. 4b again to above T_g provokes a micro-Brownian motion effect which enables the PMMA-distorted shape to return to its original shape as shown in Fig. 4c.

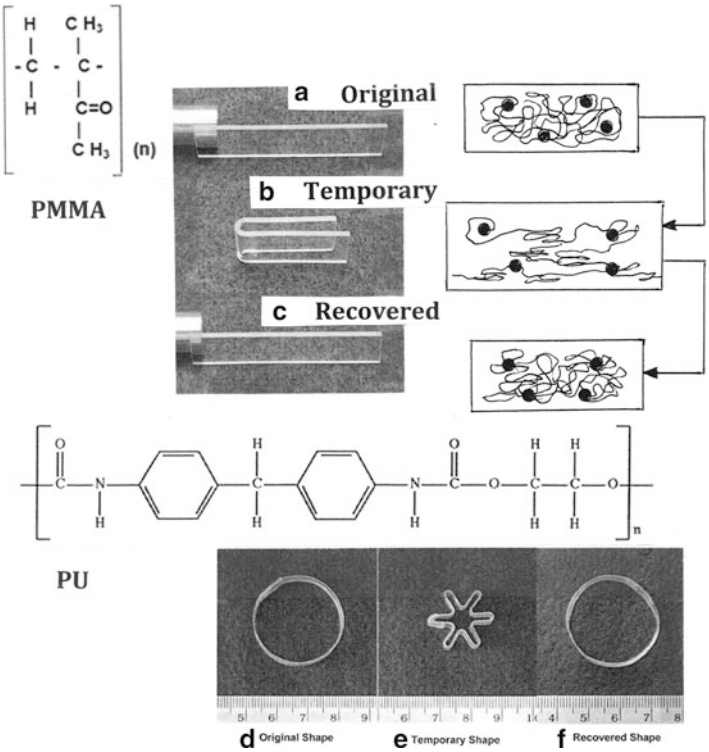


Fig. 4 Shape memory effect in thermoresponsive polymethyl methacrylate (*PMMA*) (**a**–**c**) and polyurethane (*PU*) (**d**–**f**). The corresponding molecular structures are also illustrated (Examples are adapted from Professor Wei Min Huang, Nanyang Technological University (Singapore))

A similar effect is observed in ▶ Fig. 60d–f of chapter "Chemical Forces: Nanoparticles" for polyurethane (PU), but this shape memory effect occurs because the structure is composed of two component domains: a permanent elastic component and a transitional component which is able to significantly alter its stiffness. This produces a hard/soft segment structure which responds to heating, similar to PMMA in Fig. 4a–c. This hard-to-soft transition is also characteristic of metal alloy shape memory effect.

While it can be appreciated that shape memory polymers such as those represented in Fig. 4 can be fabricated as permanent stent devices, the aliphatic polymers represented in Fig. 2 also exhibit thermoresponsive shape memory phenomena. These are simultaneously biodegradable and functional shape memory polymers and can have obvious applications for self-deployable vascular and coronary devices, especially in the endovascular treatment of aneurysm.

There are also a range of applications in medical textiles as well as related textile applications where a variety of fabrics and even fabric composites can be composed of shape memory polymer fibers (Hu 2007). Correspondingly, shape memory polymers and polymer fabrics can have a wide range of other nonmedical applications.

Early indications of shape memory effects in alloy systems were recorded in the 1930s in Au–Cd binary alloys and later in Cu–Zn alloys (Otsuka and Wayman 1998), while the prominent thermoelastic behavior of the martensitic-to-austenitic phase change was illustrated by Buehler and co-workers at the US Naval Ordnance Laboratory around 1963 (Buehler et al. 1963). This involved equiatomic Ni–Ti which was named nitinol (Nickel–Titanium Naval Ordnance Laboratory). This alloy was found to be biocompatible, and it was first used in biomedical applications in the 1970s including experimental orthopedic implants and some orthodontic applications (Yahia 2011). Ni Ti and other more recent alloys exhibit the so-called superelasticity and the shape memory effect involving only temperature or stress.

In metal alloys, the shape memory effect involves a crystal phase transition – a thermoelastic effect. The parent phase is referred to as austenite (characteristic of conventional steels). Like steels, the austenite-to-martensite phase transition is diffusionless. In Ni–Ti alloy, the austenite is bcc, while the martensite is monoclinic; in contrast to steels where austenite is fcc and martensite is bcc (or bct, body-centered tetragonal). In contrast to steel (which does not exhibit a shape memory effect), which can only exhibit recoverable plastic strain p to ~0.5 %, shape memory alloys such as Ni–Ti exhibit a maximum recoverable strain without permanent damage up to ~8 %, more than 10 times greater.

When Ni–Ti (NiTi) is cooled, the austenite changes to martensite (a hard-to-soft transition). The temperature at which this transition starts is called the martensite start temperature (M_s). Correspondingly, when the martensite NiTi is heated, it begins to change to austenite at an austenite start temperature, A_s. The temperature at which the martensite-to-austenite phase transition is complete is called the austenite finish temperature, A_f. In the reverse direction, martensite finishes at a temperature of M_f. The temperature range for the martensite-to-austenite transformation that occurs upon heating is somewhat higher than that for the austenite-to-martensite transformation upon cooling. This is characteristically a hysteresis effect as illustrated in Fig. 5a, which also shows the corresponding austenite and martensite crystal structures. The microscopic features of this shape memory effect transformation are illustrated schematically in Fig. 5b: the austenite at (1) is cooled to form twinned martensite at (2) without undergoing a shape change. Deformation (shear) of the twin plane produces the deformed martensite at (3) which characterizes a shape change. This is recovered on going from (3) to (1) in Fig. 5b. This shape recovery corresponds to A_f in Fig. 5a, and once the shape has recovered at A_f, there is no change in shape when the specimen is cooled below M_f and the shape memory can be reactivated only by deforming the martensitic material again. This is a one-time-only occurrence often referred to as a one-way shape memory effect. In practice, the hysteresis shown in Fig. 5a means that an alloy designed to be completely transformed by body temperature ($A_f < 37\ ^\circ C$) would require cooling to ~5 °C to fully retransform it into martensite at M_f.

Material in the martensite phase is soft and ductile ($E \cong 20–70$ GPa; $\sigma_y \cong 0.05–0.3$ GPa), while the austenite (parent) phase is hard and stronger ($E \cong 70–110$ GPa; σ_y 0.1–0.8 GPa). The stress-induced martensite is superelastic and can be easily

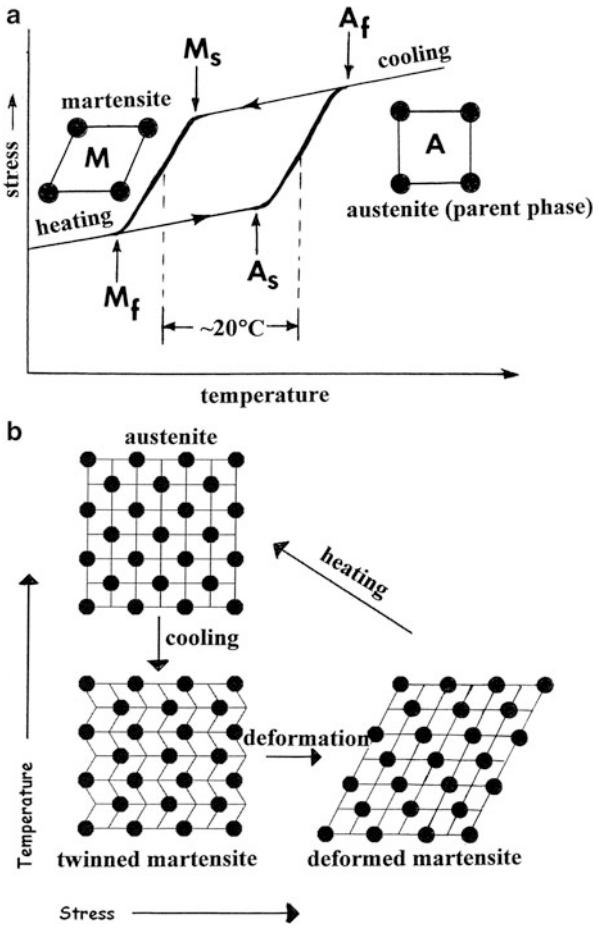


Fig. 5 Martensitic transformation hysteresis (**a**) and microscopic transformation from the austenite phase to the martensite phase characterizing the shape memory effect for NiTi (**b**). Heating either state (twinned martensite or deformed martensite) will return the original austenite structural shape. M_s martensite start temperature, M_f martensite finish temperature, A_s short of reverse transformation of martensite, A_f finish of reverse transformation of martensite

deformed or shaped. Repeated switching of the shape memory effect can lead to a shift in the transformation temperatures, M_f and A_f, a phenomenon known as functional fatigue.

In addition to NiTi (55/45), there are roughly another dozen or more different binary, ternary, and quaternary shape memory alloys: Ag–Cd (44/49 %), Cu–Zn (39/41 w/o), Ni–Ti–Cu, Ti–Pd–Ni, Ni–Fe–Zn–Al, etc. However, only NiTi is biocompatible over the long term. As illustrated in Fig. 6, NiTi (nitinol) is used in many stent applications where the stent (often cooled prior to insertion) is contained in a catheter which, in applications involving complex arterial channels, is positioned by a guide wire. Similar applications include other body conduits including the esophagus, trachea, and biliary and urinary systems. Body heat activates these devices. In some instances, inductive heating is used.

NiTi shape memory wire forms are also used as fixation devices in osteotomies in orthopedic surgery where staples are attached to each part of the bone, which,

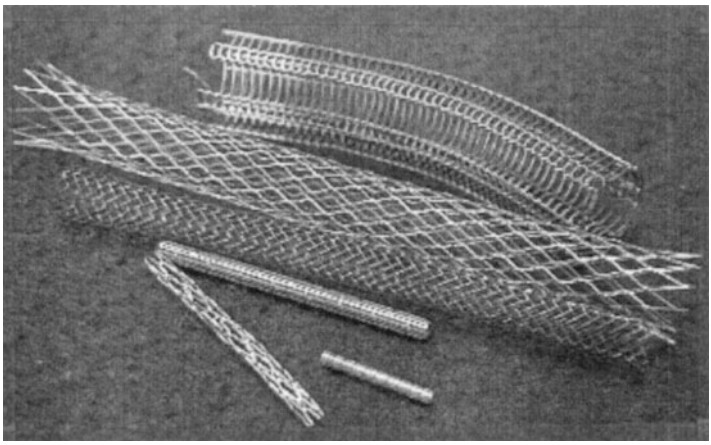


Fig. 6 NiTi commercial stents. A stent, either in conjunction with a dilation balloon or simply by shape memory effect self-expansion, can dilate or support a blocked or weakened artery or tissue conduit in the body. Millions of stents represented here are used for vascular repair or replacement annually in the United States

upon activation by body temperature, apply a constant force to pull the two pieces together to help the healing process. Similar applications are used in clinical orthodontics where these superelastic wires are used to apply a sustaining corrective force to misaligned teeth (Thompson 2000).

The variety of forms and properties (which can include the other non-biocompatible shape memory alloys noted above) of shape memory effect materials provide a wide range of non-biomedical applications. These include electrical circuit breakers, fire sprinklers, frames for eye glasses, mobile phone antennas, thermostats, and many other shape memory effect-related device structures.

Although some ceramic materials, such as certain zirconia (ZrO_2) structures, exhibit martensitic transition characteristic of shape memory effect materials, these do not exhibit prominent biomedical applications. The tetragonal-to-monoclinic crystal structure transition in ZrO_2 occurs as a martensitic phase transition induced thermally or by stress application. These are called martensitic ceramics. The transition back from monoclinic to tetragonal can occur thermoelastically, exhibiting a thermoresponsive shape memory effect.

Biomedical Applications of Auxetic Structures

On examining ▶ Fig. 16b–d in chapter "Biomimetics and Biologically Inspired Materials" and Fig. 1 as well as Fig. 6 above, it becomes readily apparent that auxetic materials and structures can have important biomedical applications either in a strict sense of a negative Poisson ratio structure or in some novel combination of biodegradable shape memory effect material. As implicit in Fig. 7, auxetic stent structures can have these combined applications, while auxetic polymer materials

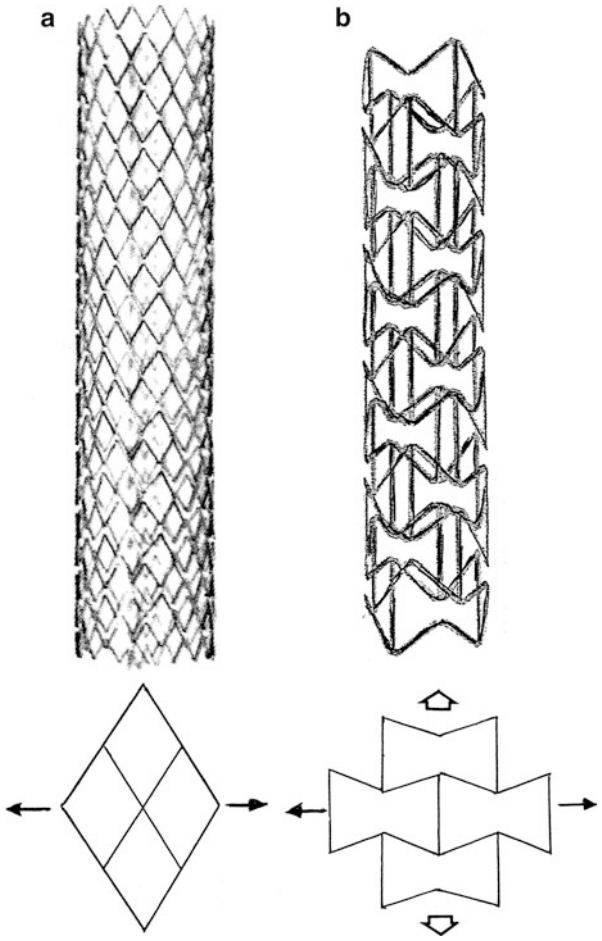


Fig. 7 Comparison of a shape memory stent (**a**) with an auxetic structure stent (**b**)

acting as artificial blood vessels will increase in wall thickness in response to pulsing blood flow and reducing the prospects for eventual rupture as a result of wall thinning (Evans and Alderson 2000). Auxetic filaments or yarns, especially polymeric materials having molecular auxetic features illustrated conceptually in ▶ Fig. 17 in chapter "Biomimetics and Biologically Inspired Materials", can form intelligent (or smart) fabrics to reduce clothing pressure or to create special bandages or wound pressure pads.

Carrie Bakker (2011) recently examined a novel possibility of using a negative Poisson ratio artificial intervertebral disc to replace degenerated spinal discs while allowing some range of motion that the natural disc allows and preventing interference with the surrounding nerves due to lack of bulging when compressed. One mechanism involves the compression of medium-density polyurethane foam to create a negative Poisson ratio foam. As illustrated in Fig. 4d–f, polyurethane can also have a thermoresponsive shape memory effect.

References

Baker CE (2011) Auxetic spinal implants: consideration of negative Poisson's ratio in the design of artificial intravertebral disc. MS Thesis, University of Toledo, Bioengr

Bonfield W, Grynpas M, Tully A, Bowman J, Abram J (1981) Hydroxyapatite reinforced polyethylene – a mechanically compatible implant material for bone replacement. Biomaterials 2(3):185

Buehler WJ, Gilfrich JW, Wiley RC (1963) Effects of low temperature phase changes on the mechanical properties of alloys near composition TiNi. J Appl Phys 34:475–480

Evans KE, Alderson A (2000) Auxetic materials: functional materials and structures from lateral thinking. Adv Mater 12:617–628

Geiger M, Li RH, Friess W (2003) Collagen sponges for bone regeneration with rhBMP-2. Adv Drug Deliv Rev 55:1613–1629

Hornberger H, Virtanen S, Boccaccini AR (2012) Biomedical coatings on magnesium alloys – a review. Acta Biomater 8:2442–2455

Hu J (2007) Shape memory polymers and textiles. Woodhead Publishing, London

Huang WM, Ding Z, Wang CC, Wei J, Zhao Y, Purnawali H (2010) Shape memory materials. Mater Today 13:54–61

Hufmacher DW, Sittinger M, Risbud MV (2004) Scaffold-based tissue engineering: rationale for computer-aided design and solid free-form fabrication systems. Trends Biotechnol 22:354–362

Huse EC (1878) A new ligature? Chicago Med J Exam 172:2

Kim Y-H, Lee B-T (2011) Novel approach to the fabrication of an artificial small bone using a combination of sponge replica and electrospinning methods. Sci Technol Adv Mater 12:1–7

Morevej M, Mantovani D (2011) Biodegradable metals for cardiovascular stent application: interests and opportunities. Int J Mol Sci 12(7):4250–4270

Ohya Y, Takahashi A, Nagahama K (2011) Biodegradable polymeric assemblies for biomedical materials. Adv Polym Sci 247:65–114

Otsuka K, Wayman CM (1998) Shape memory materials. Cambridge University Press, Cambridge, UK

Peuster M, Wohlsein P, Brugmann M, Ehlerding M, Seidler K, Fink C, Brauer H, Fischer A, Haasdorf GA (2001) A novel approach to temporary stenting: degradable cardiovascular stents produced from corrodible metal – results 6–18 months after implantation into New Zealand white rabbits. Heart 86:563–569, 370

Razavi M, Fethi M, Savabi O, Boroni M (2012) A review of degradation properties of Mg based biodegradable implants. Res Rev Mater Sci Chem 1(1):15–58

Shen D, Fang L, Chen X, Tang Y (2011) Structure and properties of polyacrylic acid modified hydroxyapatite/liquid crystal polymer composite. J Reinf Plast Compos 30(13):1155–1163

Thompson SA (2000) An overview of nickel-titanium alloys used in dentistry. Int Endod J 33:297–310

Wahl DA, Czernuszka JT (2006) Collagen-hydroxyapatite composites for hard tissue repair. Eur Cells Mater 11:43–56

Wang M, Bonfield W (2001) Chemically coupled hydroxyapatite-polyethylene composites: structure and properties. Biomaterials 22(11):1311–1320

Witte F (2010) The history of biodegradable magnesium implants: a review. Acta Biomater 6:1680–1692

Xu Z, Ikoma T, Yoshioka T, Tagaya M, Motozuka S, Matsumoto R, Vemura T, Tanaka J (2012) Effect of glutaraldehyde on properties of membranes prepared from fish scale collagen. MRS Proc 1418:2012–2396. doi:10.1557/op1

Yahia L (ed) (2011) Shape memory implants. Springer, New York

Implant Materials and Structures

Contents

Abstract

In contrast to biodegradable implant materials and appliances, many implant metals and alloys are permanent: teeth, cranial-facial, orthopedic appliances for knee and hip replacement, rods, screws, etc. The design and fabrication of biomechanically compatible orthopedic (bone) implants will be discussed, especially with regard to reducing or eliminating stress shielding through controlled density and porosity to reduce stiffness (or Young's modulus).

Introduction

The implications of implant materials, especially metals or alloys are certainly implicit in ▶ Fig. 3 (chapter "Strategies for Bone Replacement and Tissue Augmentation"), in spite of the biodegradable features in ▶ Fig. 3 of chapter "Strategies for Bone Replacement and Tissue Augmentation". Here an appreciation for more permanent metallic implants and related biomaterials appliances will be developed.

The setting of bones damaged by fracture or disease dates to ancient times. Early attempts to correct hip and hip joint deformities involved prosthetic hip replacements (especially femoral heads) carved from ivory and using pumice or plaster of Paris ($CaSO_4 \cdot 2H_2O$) to secure the implant occurred around 1890 (Walker 1978). Archaeologists have found evidence that occupants of the region which is now Honduras, around 1000 B.C., developed a way to use tooth-shaped stones as dental implants. Lead was also used for dental implants in Europe in the late 1800s.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_32

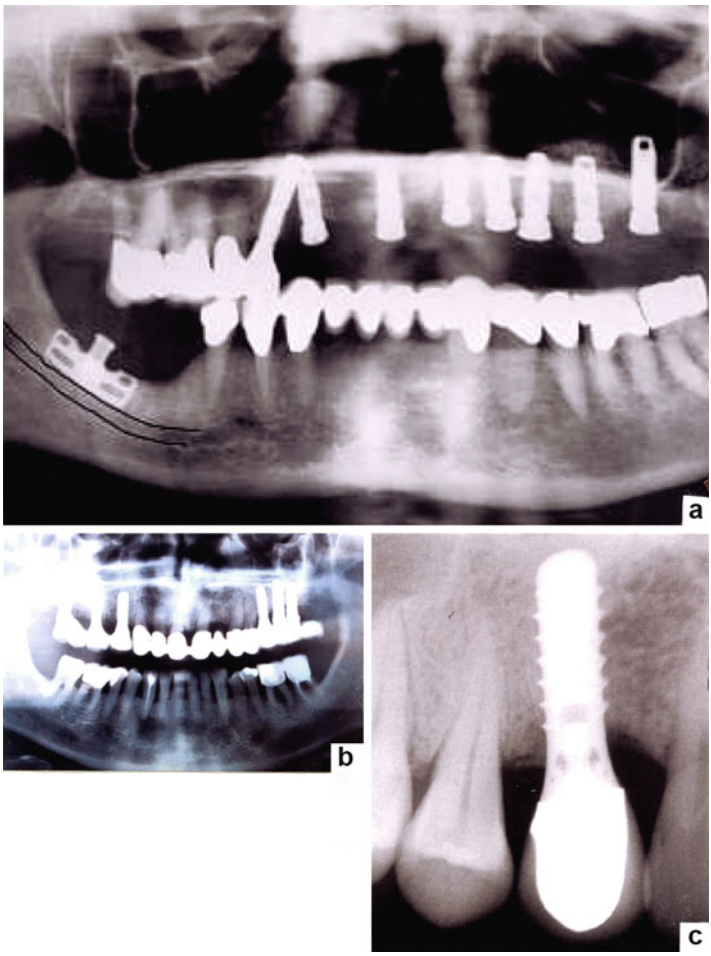


Fig. 1 Examples of dental implants and appliances: X-ray images in (**a**) and (**b**). (**c**) shows an X-ray image of osseointegrated tooth-screw implant (Courtesy of Bing images)

Metal inserts such as plates and screws to secure bones became possible after the discovery of antiseptics by Joseph Lister and the development of X-radiography around 1895. In 1886, D. H. Hansmann, a German surgeon, was among the first to use metal plates of vanadium steel to secure bone fractures, but this material was not biocompatible. With the invention of biocompatible stainless steels around 1926, these materials became the implant material (alloy) of choice until around the 1970s when stainless steel corrosion was recognized as a consequence of deformation-induced martensite formation either as a processing feature or as a consequence of direct deformation of biomedical appliances. Cobalt-based alloys were introduced as implant biomaterials in 1936 while Ti alloys were introduced in the 1950s. Cobalt-chrome and Ti alloys continue to be major implant materials,

Fig. 2 X-ray images showing total femur fracture at *arrow* in (**a**) and healed femur with intramedullary stem insert (**b**) after 18 months (Courtesy of Prof, A. M. Ferracini)

especially permanent implant materials, even today (2014), including complete knee and hip replacement structures, pins, screws, rods, plates, etc. used in bone repair and augmentation as implicit in ▶ Fig. 3 in chapter "Strategies for Bone Replacement and Tissue Augmentation". All of these alloy devices are biocompatible.

Dental Implants

Teeth have become a popular implant, especially Ti and Ti alloys such as Ti-6Al-4V. Equally popular are screw-type implants. Examples are shown in X-radiographic images in Fig. 1. Because Ti and its alloys produce a TiO_2 surface oxide with some porosity, these implants encourage bone cell ingrowth at the implant surface. In addition, TiO_2 can exhibit antibacterial properties. However, Ti alloy implants in the bone exhibit stress shielding since Young's modulus normally averages 110 GPa in contrast to cortical bone with a Young's modulus of 15–30 GPa.

Although pure Ti dental implants are only half as strong as Ti-6Al-4V implants, recent development of nano-Ti dental implants has, as a consequence of the nano-sized grain structure, improved the strength to roughly that of Ti-6Al-4V (~l10 GPa) in contrast to pure Ti with strengths ranging from 0.2 to 0.8 GPa (Murr et al. 2012a). Ceramic implants have also involved Al_2O_3 with a Young's modulus of 0.35 GPa and ZrO_2 with a Young's modulus of 0.2 GPa, more compatible with that of cancellous bone than either Ti or Ti alloys. Ceramic implants are also more biocompatible than metals, but are mechanically less suitable.

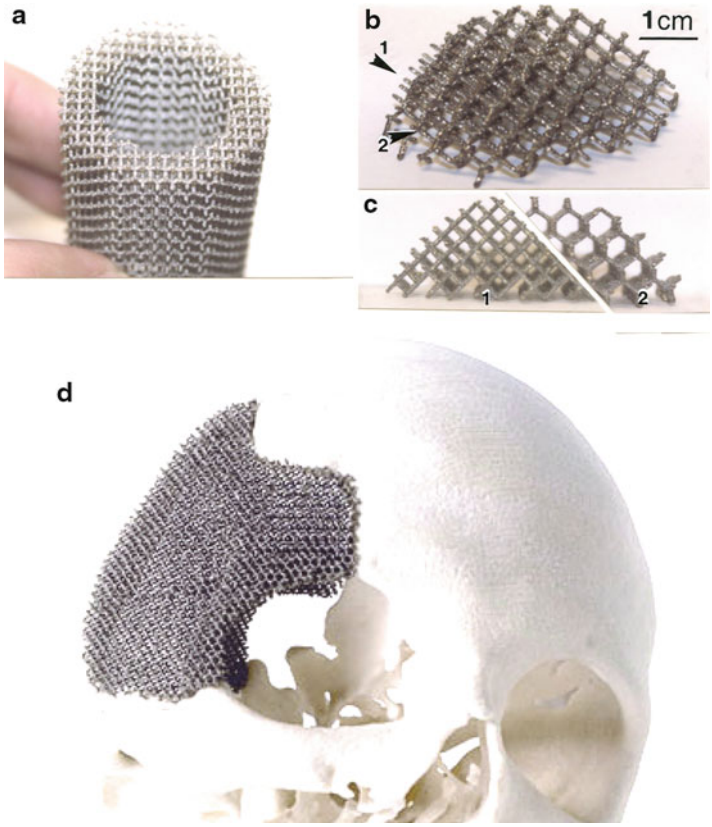


Fig. 3 Reticulated (Ti-6Al-4V) mesh prototypes fabricated by electron beam melting: (**a**) to (**c**). *1* and *2* in (**b**) and (**c**) indicate mesh geometry differences at 45° relative to one another. (**d**) shows a skull-bone section replacement implant (After Murr et al. 2012a)

The concept of plate-and-screw implant configurations to stabilize bone breaks or compromised bone segments is illustrated in ▶ Fig. 3 of chapter "Strategies for Bone Replacement and Tissue Augmentation", which can conceptually represent more permanent metal or alloy implant components. Ti alloy rods, such as Ti-6Al-4V, are also employed in major fractures of the femur as illustrated typically in Fig. 2. Normally these intramedullary rod implants are permanent, and as illustrated in Fig. 2, they replace the interior, soft, trabecular bone.

Metal or alloy (especially Ti-6Al-4V) plate configurations as cranial-facial or maxillofacial implants have provided skull-bone replacements for several decades, but like other solid implant configurations (Fig. 2), they do not allow for bone ingrowth, although the lack of stress loading eliminates bone remodeling in response to stress shielding (difference in elastic moduli) which occurs for implanted rods or other appliances characteristic of Fig. 2. The lack of bone

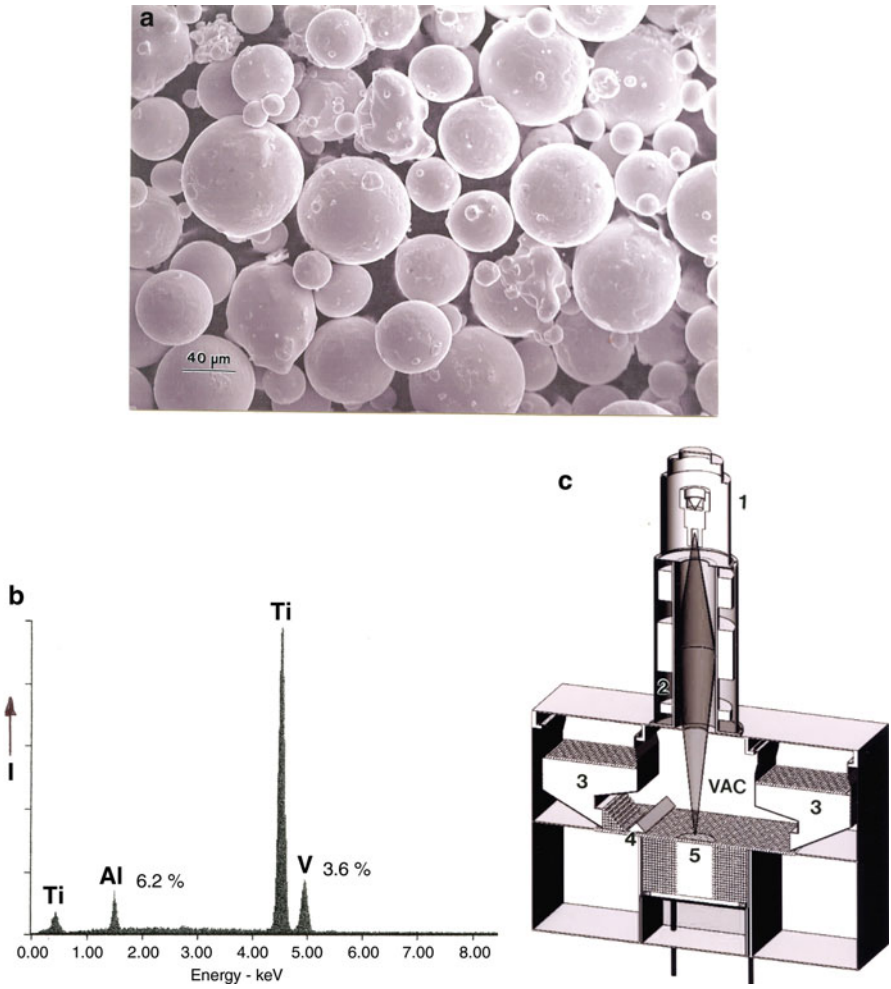


Fig. 4 Electron beam melting (EBM) of pre-alloyed Ti6Al-4V powder: (**a**) and (**b**). The EBM system is shown schematically in (**c**), consisting of an electron gun (*1*), focusing lenses and beam scan coils (*2*), powder cassettes at (*3*), layer raking at (*4*), and product building at (*5*)

ingrowth at the implanted plate margins essentially leaves the insert freely supported only by skin and surrounding tissue. Although generally not in practical use, the ability to manufacture patient-specific or bone-replacement-specific porous or open-cellular implants of Ti-6Al-4V by additive manufacturing can eliminate these shortcomings as illustrated in Fig. 3d. Figure 3a–c shows examples of reticulated mesh prototypes fabricated by electron beam melting (EBM) illustrated in Fig. 4. This additive manufacturing technology, to be discussed in more detail in Part XI, chapter "► Laser and Electron Beam Melting Technologies", utilizes build

Fig. 5 X-ray image showing contemporary (commercial) total hip replacement. The femoral stem is indicated by F_s (Adapted from Murr et al. (2012a))

units or geometrical unit cells embedded in computer-aided design (CAD) software to selectively melt powder layers by CAD-directed electron beam scanning and melting.

The most common permanent implant strategies involve partial and total hip and knee replacements, which in the aggregate now number in millions annually worldwide. Figure 5 illustrates a total hip replacement, while Fig. 6 shows the replacement appliance components. The acetabular cup is usually Ti-6Al-4V and in some instances, as shown in Fig. 6 (at A), is fabricated by electron beam melting technology to create a porous contacting surface to promote hip-bone cell ingrowth. Bone screw holes for attachment are also shown. The liner shown at L in Fig. 6 is usually made of highly cross-linked polyethylene (HCLPE) (or ultrahigh molecular weight polyethylene (UHWWPE)), while the femoral head (FH in Fig. 6) is usually Ti-6Al-4V or harder cobalt-chrome alloy (Co-29 % Cr, 6 % Mo, 0.7 % Si, 0.25 Ni, 0.22 % C, in weight percent). Acetabular cups as shown in Fig. 7 at Pc have been manufactured commercially by EBM. Femoral heads have also been made from hard ceramic such as Al_2O_3, while the femoral stem (F_s) has been made from Ti, Ti-6Al-4V, Co-Cr alloy, and 316 stainless steel. Ti-6Al-4V is a common commercial choice. Femoral stem designs can vary as illustrated in Fig. 7 where the middle design shows a porous coating (also a component of the femoral stem shown at P in Fig. 6). Such coatings can involve sintered powder beads or open-cellular structures such as tantalum or other biocompatible, porous structures as illustrated in Fig. 8. Plasma-sprayed hydroxyapatite (HA) coatings have also been successfully applied to Ti and Ti-6Al-4V to increase bone integration of both orthopedic and dental prostheses (Brunski et al. 2000).

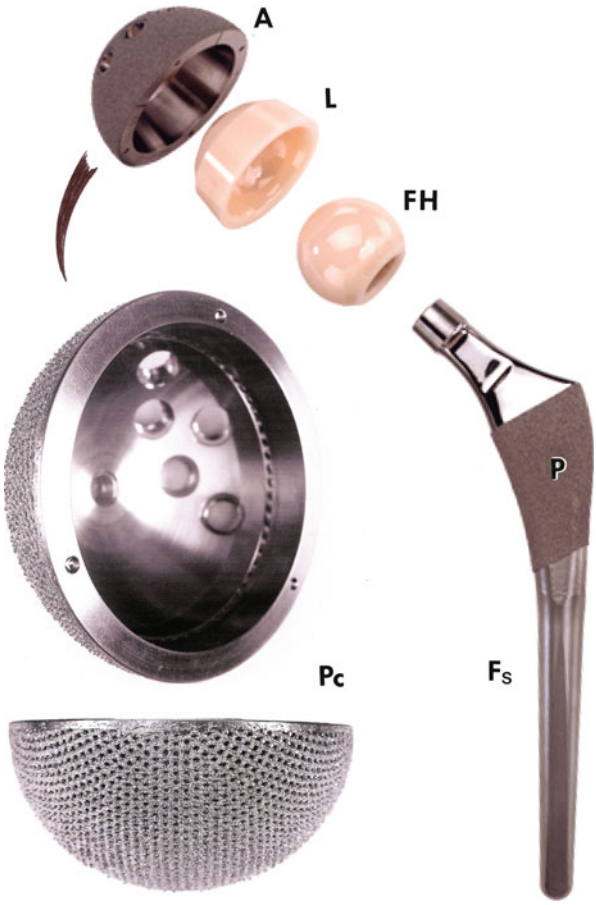


Fig. 6 Total hip appliance components. *A*: acetabular cup. The *arrow* shows an EBM-fabricated (Ti-6Al-4V) component having a porous (mesh) coating or surface mesh (P_c). *L* is the polyethylene liner. *FH* is the femoral head. F_s is the femoral head. F_s denotes the femoral stem which has a porous coating section denoted *P*

Figure 9 shows the microstructure typical for wrought (forged) Ti-6Al-4V alloy characteristic of solid femoral stems as illustrated in Fig. 7. Figure 9a shows acicular alpha (hcp: a = 0.3 nm, c = 0.47 nm) grains which are surrounded by intergranular beta-phase (bcc: a = 0.33 nm). Figure 9b, c shows increasing TEM magnifications representing a region corresponding to the rectangle (arrow) in Fig. 9a. Similar microstructures are observed for solid Ti-6Al-4V components fabricated by electron beam melting (EBM) as illustrated in Figs. 10 and 11. In Fig. 10, which is characteristic of a solid cylinder fabricated by EBM, the heat flow creates slightly larger α-grains (resembling a Widmanstätten structure similar to ▶ Fig. 4b in chapter "Volume Defects: 3D Imperfections in Crystals") near the upper portion of the component. In Fig. 11a, b, corresponding to a microstructure similar to Fig. 10b, a heavy dislocation density contributes to a higher hardness than a relatively dislocation-free portion characteristic of Fig. 10a as shown in Fig. 11c, d.

Fig. 7 Examples of femoral stem designs. The middle example has a porous surface coating (Courtesy of DiSanto, Inc)

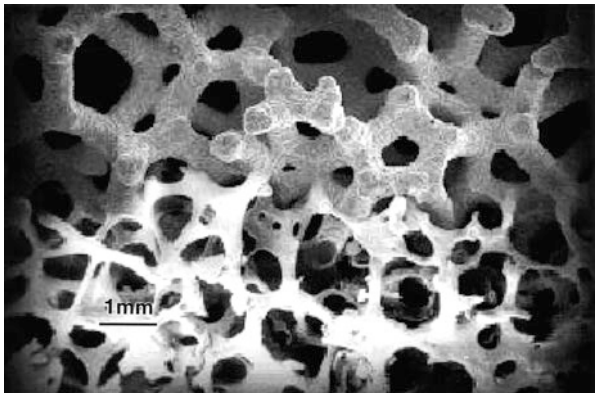


Fig. 8 Open-cellular tantalum (Trabecular Metal®) developed by Zimmer, Inc., for implant surface porosity on commercial implant appliances (Courtesy of Zimmer Holdings, Inc)

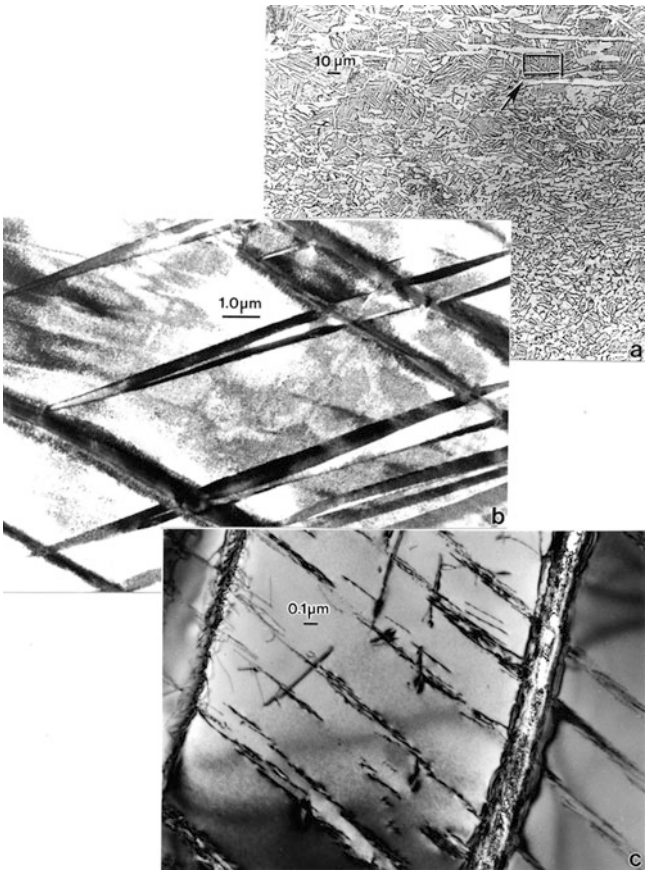


Fig. 9 Wrought Ti-6Al-4V microstructures. (**a**) Optical microscopy view. (**b**) and (**c**) show TEM bright-field images. In (**a**) black β-phase surrounds α-grains

There are several important issues to consider with respect to permanent implants. First is the necessity to assure their integrity or fixation, and in many cases they are cemented into the bone, often using acrylic polymer material such as PMMA. Porous or open-cellular surface structures as illustrated in Figs. 6 and 8 promote bone cell ingrowth and provide permanent fixation. These porous structures can also lower Young's modulus next to the bone and reduce stress shielding. We will discuss an engineering design strategy to address this issue later. It can be appreciated in Fig. 7 that commercial implants are manufactured in selected sizes while human bones come in a great variety of sizes. Solid stems also eliminate the natural vascularization, and if porous metal implants could be fabricated with varying, functional porosities or open-cellular structures, it may be possible to grow vascular regimes to allow for some blood flow. In this regard, Karageorgiou and Kaplan (2005) have recommended that pore sizes of over 300 μm favor direct osteogenesis and allow potential vascularization. Similarly, the general

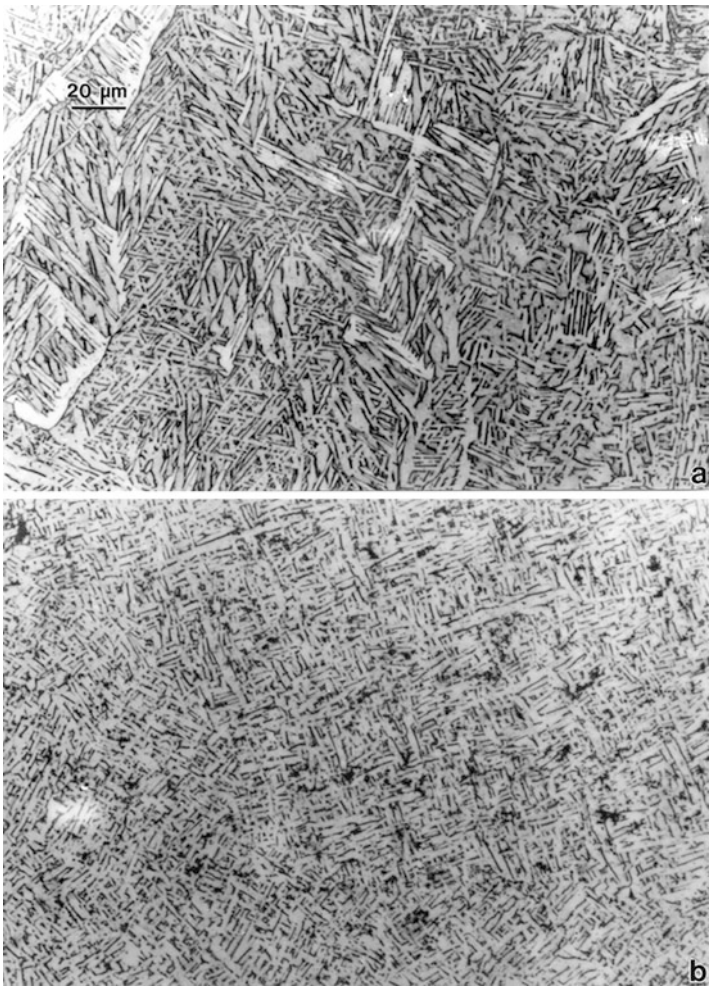


Fig. 10 Alpha-phase microstructure in EMB-fabricated Ti-6Al4V cylindrical component. (**a**) represents *top* section, while (**b**) represents cooler *bottom* section with smaller grain (phase) size

consensus has been that a minimum pore size in the range of 100–500 µm promotes the formation of fibrous tissues in bone cell ingrowth (Freyman et al. 2001; Ryan et al. 2006). Bone cell ingrowth does not generally occur for pore sizes less than 100 µm.

Murr and coworkers (Murr et al. 2010, 2011, 2012a, b) have recently promoted the prospects for creating novel, next-generation orthopedic hip and knee implants consisting of open-cellular (porous) monoliths fabricated by EBM. These can be

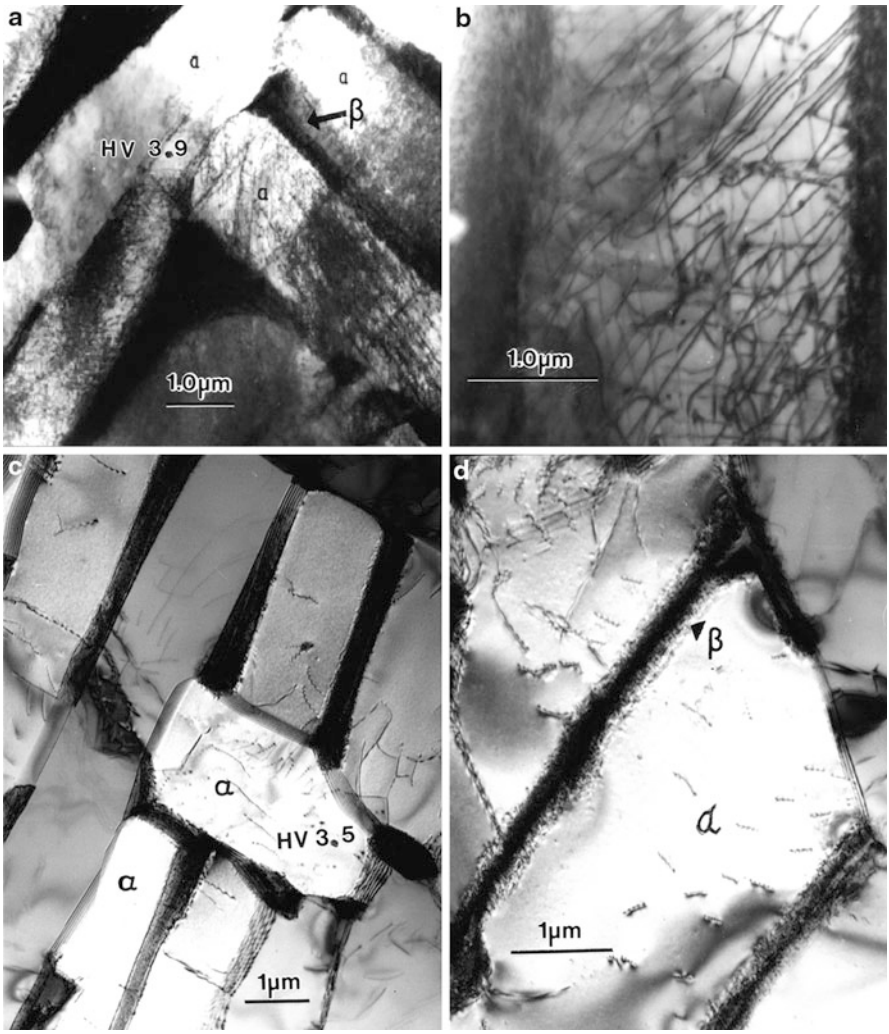


Fig. 11 Transmission electron microscope images showing phase and dislocation structures in EBM-fabricated Ti-6Al-4V components. (**a**) and (**b**) show a dislocation substructure while (**c**) and (**d**) show a relatively dislocation-free structure. Vickers hardness is denoted *HV*

fabricated from CAD models created from CT or micro-CT scans of particular implant patients to create patient-specific appliances. Figure 12 illustrates this concept for intramedullary (femoral) rods. In Fig. 12a, the rod consists of an outer, smaller-size foam, while the inner portion is a larger foam size (pore size). In Fig. 12b, the outer rod structure is a special build unit corresponding to a reticulated mesh while the inner structure is a foam as in Fig. 12a. Figure 13

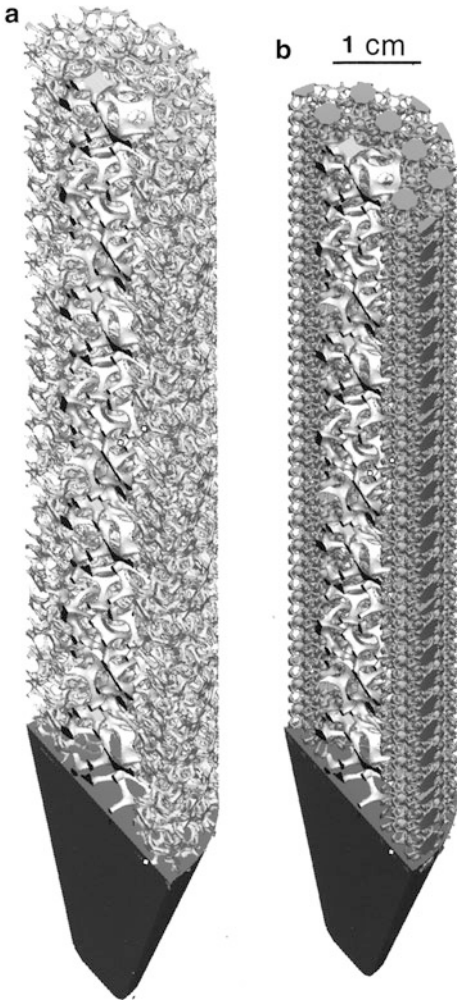


Fig. 12 CAD models for open-cellular intramedullary rod monoliths. (**a**) Higher-density foam surrounding lower-density interior. (**b**) Reticulated higher-density mesh surrounding lower-density foam interior (Adapted from Murr et al. (2012b))

shows more detailed views of these models: Fig. 13a–c corresponds to Fig. 13a, while Fig. 13d corresponds to Fig. 11b. Figure 14 shows a Ti-6Al-4V stem corresponding to Fig. 12a, fabricated by EBM and placed in a femur bone.

Open-cellular foams and reticulated mesh structures are characterized by thin strut or cell walls which when fabricated by EBM as illustrated in Fig. 14 cool more rapidly than bulk components. In Ti-6Al-4V this produces an α'-martensite structure as illustrated in Fig. 15. This microstructure, in contrast to the normal α-phase, acicular grain structure shown in Fig. 10, is roughly 25 % harder and characteristically stronger.

Knee replacement biomaterials are essentially the same as those used in permanent hip replacement, and their function is similar. As illustrated in Fig. 16,

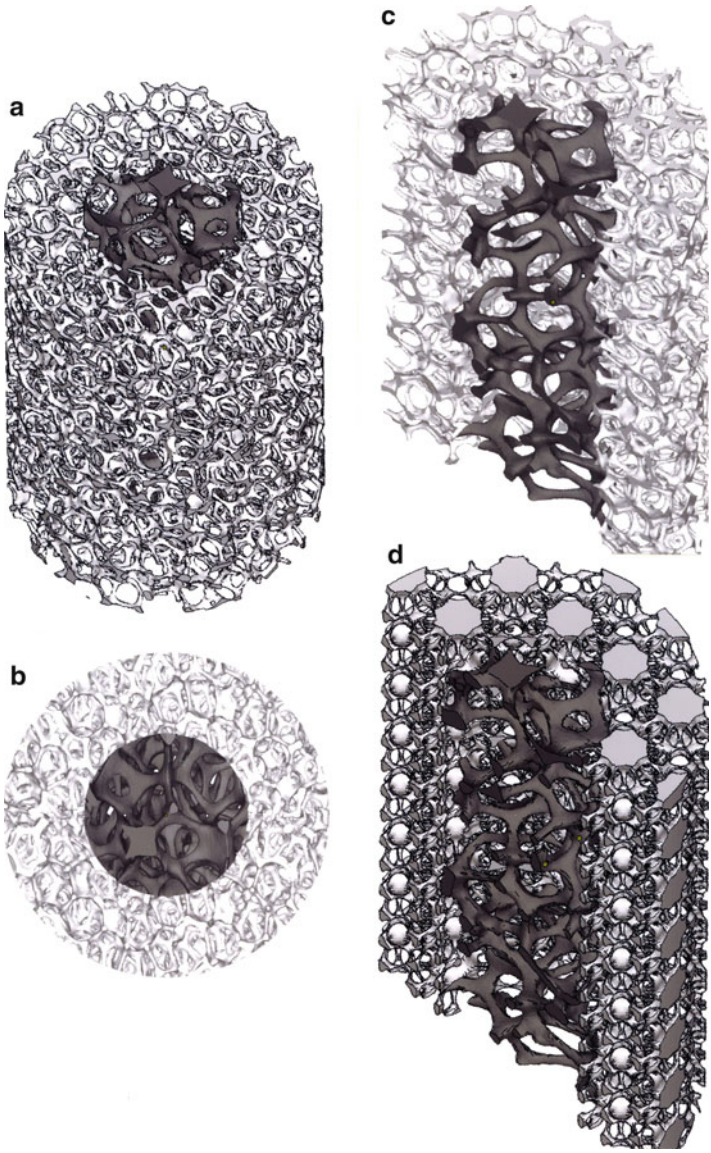


Fig. 13 CAD models representing open-cellular structures shown in Fig. 13, although (**a**) and (**b**) show different foam density models. (**c**) shows an axial view of (**a**), while (**d**) corresponds to Fig. 13b

a femoral component, usually hard Co-Cr alloy, rides on a highly cross-linked (UHMWPE) polyethylene pad fixed to the tibial table inserted into the tibia and usually made of Ti-6Al-4V. The polyethylene insert, not visible in the X-ray image in Fig. 16a, essentially replaces the damaged and often completely absent meniscus,

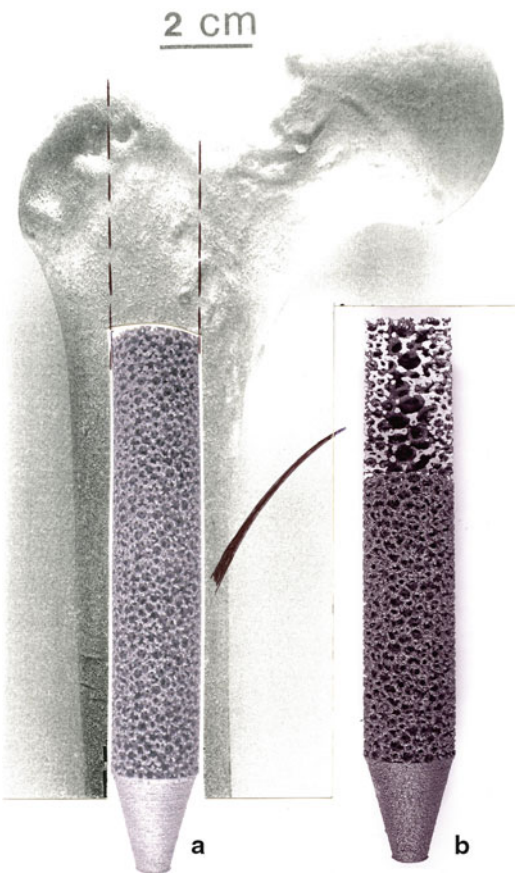


Fig. 14 Upper femur with simulated rod insert of Ti-6Al-4V corresponding to ▶ Fig. 75a of chapter "Chemical Forces: Nanoparticles" and fabricated by EBM (**a**). (**b**) shows a cutaway section of the porous rod in (**a**) (Adapted from Murr et al. (2012b))

which, although not visible in the opposite knee in Fig. 16b, has deteriorated to a thin membrane separating the femur from the tibia. The left knee in Fig. 16b was subsequently implanted with a total knee replacement as illustrated in Fig. 17a, utilizing an 11.5 cm incision shown in Fig. 17b for preparation and insertion of the appliances.

As in the case of hip replacement, acrylic cements are used to secure the implants although porous surface preparations for bone ingrowth are common. In addition, cements are often mixed with antibiotics such as gentamicin and norfloxacin to guard against infection (Baleari et al. 2003; Vasilev et al. 2009). Silver nanoparticles (▶ Fig. 10b in chapter "Examples of Materials Science and Engineering in Antiquity") have also been incorporated into polymeric cements (Camacho 2012) and can be embedded in open-cellular, porous structures as well (Murr et al. 2012a).

The weak link in knee implant systems in particular involves degradation of the polyethylene pad which can crack by impact and wear by granulation.

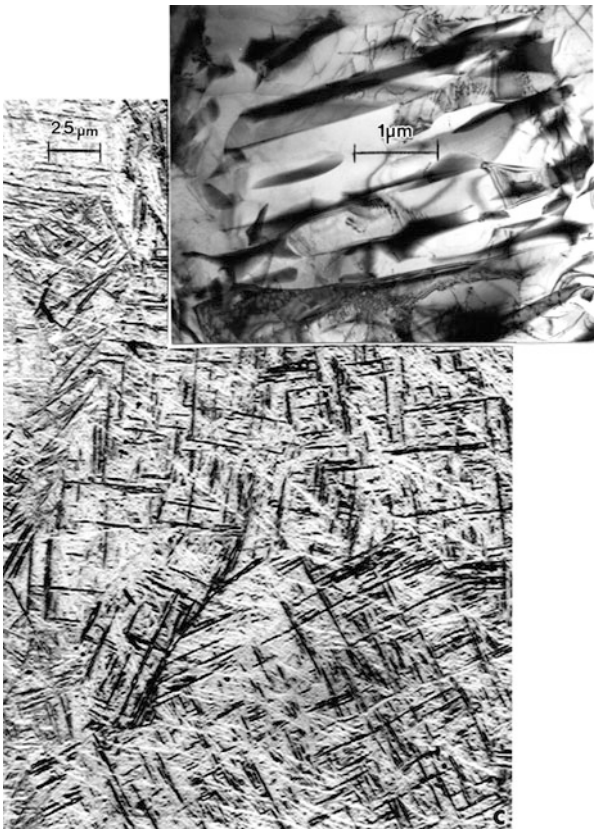


Fig. 15 Optical microscope image and TEM image insert showing α′-martensite in Ti-6Al-4V open-cellular foam structure as is Fig. 15 fabricated by EBM. Compare with the solid Ti-6Al-4V α-phase structure in Fig. 11 and Widmanstätten structure in ▶ Fig. 4b of chapter "Volume Defects: 3D Imperfections in Crystals"

This can include incorporation of arthritic debris which is often the cause of initial tissue and ligature degradation. Recent research has involved the incorporation of carbon nanotubes within the polyethylene, which has demonstrated wear life (or wear resistance) more than 86 % greater than unreinforced, ultrahigh molecular weight polyethylene (Camacho 2012). Fatigue is also an important implant consideration along with surface preparation of the tibial appliance which rides on the polyethylene pad. Scratches of any kind can damage the polyethylene surface as well.

Figure 18 illustrates an experimental Co-Cr tibial component fabricated by EBM which has an open-cellular interior surface to promote tibial bone ingrowth. The novel microstructure produced by this processing is illustrated in Figs. 19 and 20 which show columns of $Cr_{23}C_6$ cubic precipitates (a = 10.6 Å) especially prominent in the TEM image of Fig. 20. Table 1 compares EBM-fabricated solid Ti-6Al-4V and Co-Cr components along with commercial, wrought product nominal properties. Nominal stiffnesses for Ti-6Al-4V and Co-26Cr-6Mo alloys are 110 GPa and 210 GPa, respectively.

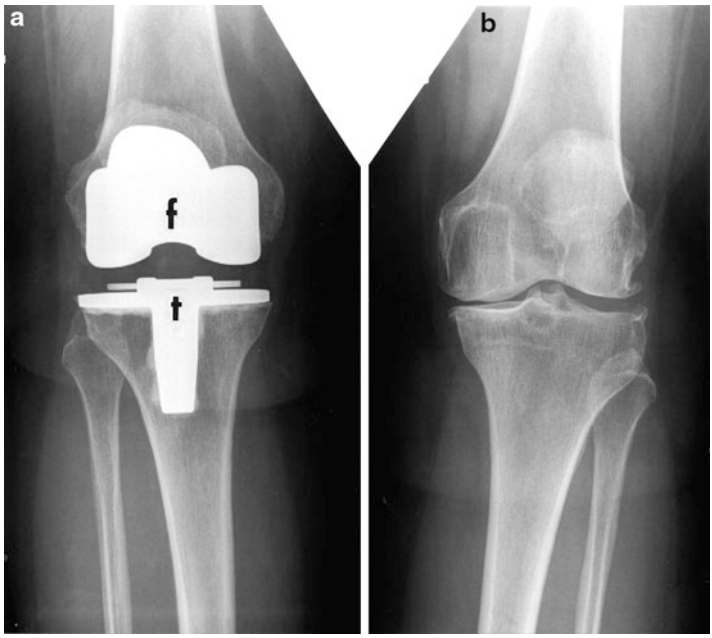


Fig. 16 X-ray images illustrating a total knee replacement in right knee (**a**) and compromised meniscus in left knee (**b**). *f* denotes femoral appliance and *t* denotes tibial appliance (Courtesy of Patricia Murr)

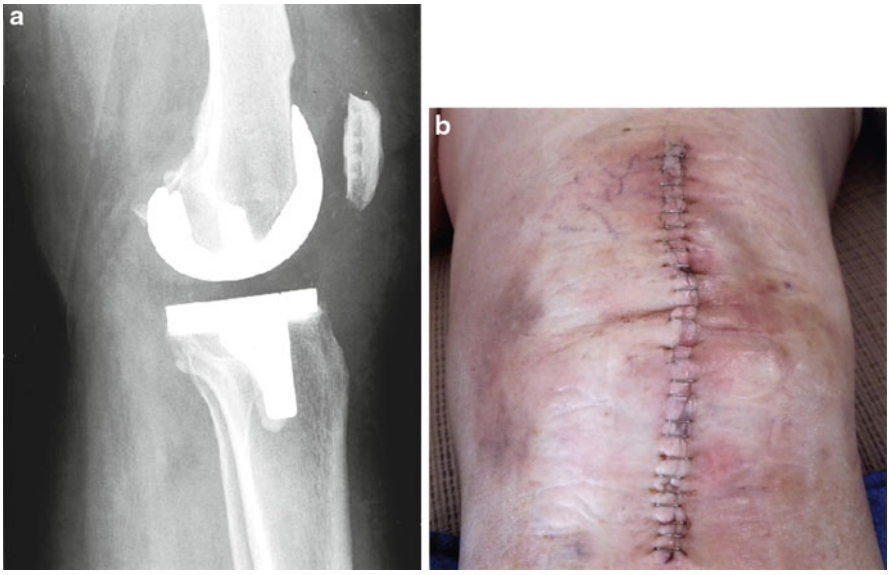


Fig. 17 X-ray image of total knee replacement in left knee of ▶ Fig. 79b of chapter "Chemical Forces: Nanoparticles" (**a**). (**b**) shows the replacement incision for (**a**) (Courtesy of Patricia Murr)

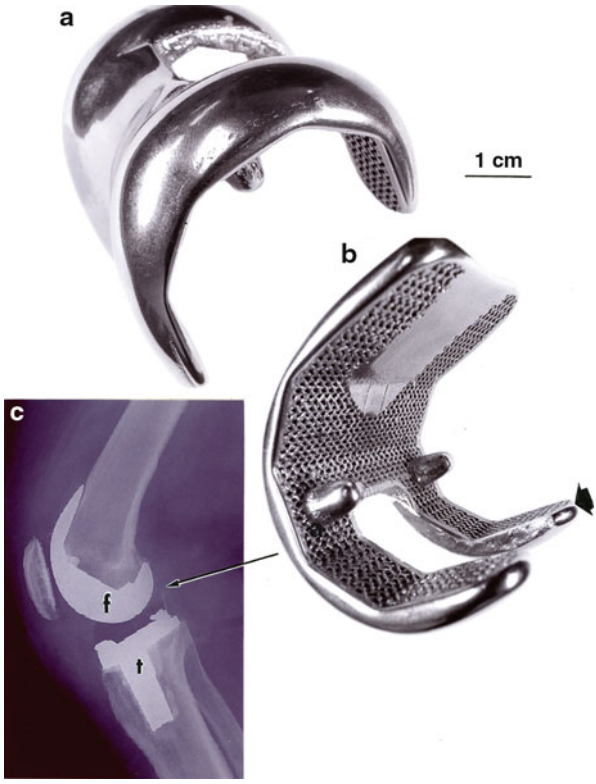


Fig. 18 Co-Cr EBM prototype femoral appliance showing outside (**a**) and inside (**b**) surface structures. (**c**) shows the position of the femoral appliance in an X-ray image of the total knee replacement implants

Figure 21 shows a Co-Cr reticulated mesh structure, while Fig. 22 shows this structure embedded in the CAD model and actual fabrication of a Ti-6Al-4V tibial implant for an exaggerated stem prototype, using EBM. Figure 23a shows an SEM image for the Ti-6Al-4V reticulated mesh strut structure for comparison with that for a Co-Cr reticulated mesh strut structure in Fig. 21a. Figure 24 compares the microstructure for reticulated Ti-6Al-4V (Fig. 24a) and Co-Cr (Fig. 24b) mesh structures which correspond to those shown in Figs. 15 and 19, respectively.

On examining Figs. 6, 12, 13, 14, 18a, b, 22 in retrospect, it should be apparent that as a consequence of the ability of additive manufacturing (especially using EBM) to fabricate specific, functional, monolithic implants and components, it should be possible to design patient-specific and biomechanically compatible appliances. This is especially true for the selection of Young's modulus which is related to the density or specific density (open-cellular structure density, ρ, divided by the solid density, ρ_s (ρ/ρ_s)). The elastic (Young's) modulus for an open-cellular foam based on a truncated cuboctahedron strut unit can be explicitly calculated using the simple principle of Egli (1972):

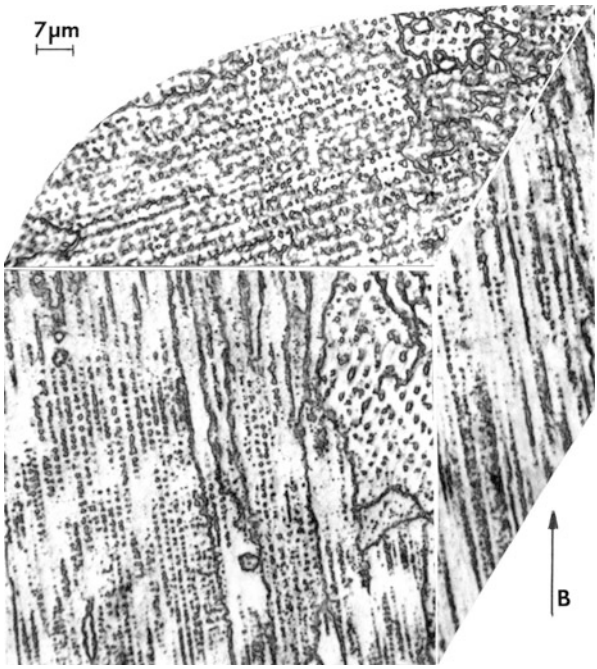


Fig. 19 3D image composite (optical microscopy) for EBM-fabricated Co-Cr solid cylinder characteristic of the solid appliance in Fig. 19a, b. *B* denotes the build direction

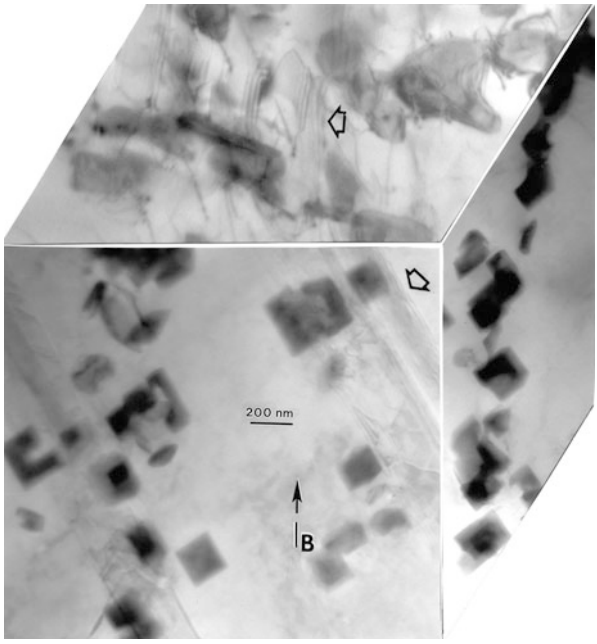


Fig. 20 TEM bright-field image 3D composition corresponding to Fig. 19 showing $Cr_{23}C_6$ carbide precipitates. *B* denotes the build direction as in Fig. 19. Open *arrows* show stacking fault images

Table 1 Comparison of mechanical properties for solid Ti-6Al-4V and Co-26Cr-6Mo alloys

Material	Microstructure	HV (GPa)	HRC	YS (GPa)	UTS (GPa)	Elongation (%)
Wrought Ti-6Al-4V	Coarse acicular α-phase	3.8	48	1.17	1.23	12
EBM (z) Ti-6Al-4V	Acicular α-phase	3.6	38	1.03	1.11	13
EBM (x, y) Ti-6Al-4V	α′-martensite + α-phase	4.1	41	1.10	1.11	11
ASTM Grade 5 (nominal) Ti-6Al-4V	Acicular α-phase plates	3.4	37	0.9	1.00	15
ASTM-F75 (wrought) Co-26Cr-6Mo	Equiaxed grains with annealing twins	2.8	30	0.45	0.66	10
EBM (z) Co-26Cr-6Mo	Columnar precipitates ($Cr_{23}C_6$)	4.6	48	0.51	1.45	3.6
EBM (x, y) Co-26Cr-6Mo	Columnar precipitates ($Cr_{23}C_6$)	–	–	0.51	1.45	2.7
EBM (z) + HIP Co-26Cr-6Mo	Equiaxed fcc grains with annealing twins	3.6	31	0.60	1.15	32
EBM (x, y) + HIP Co-26Cr-6Mo	Equiaxed fcc grains with annealing twins	–	–	0.63	0.99	20

ASTM-F75 standard
EBM (z) represents samples prepared from a reference plane perpendicular to the EBM build direction (z-axis as in Fig. 4)
EBM (x, y) represents samples prepared or built perpendicular to the z-axis (in the x- or y-axis direction)
Microstructure is represented by optical metallography images (Adapted from Murr et al. (2012b))

$$E/E_s \propto (\rho/\rho/_s)^n \tag{1}$$

and considering the Castigliano method (Prawoto 2012)

$$\delta = \frac{\partial U}{\partial P} = \sum_i^n \int_Z \left(\frac{M_i}{EI}\right)\left(\frac{\partial M_i}{\partial P}\right) dz \tag{2}$$

where δ is the deflection of a thin cantilever beam loaded at its end with a load, P, U is the strain energy, M_i is the bending moment exerted at each (ith) beam, E is Young's modulus, I is the area moment of inertia, and n is the number of beams in the structure. Applying Eq. 2 to the truncated cuboctahedron representing a foam cell (or CAD unit cell), Young's modulus was calculated to be (Prawoto 2012):

$$E/E_s = 0.9\,(\rho/\rho_s)^2 \tag{3}$$

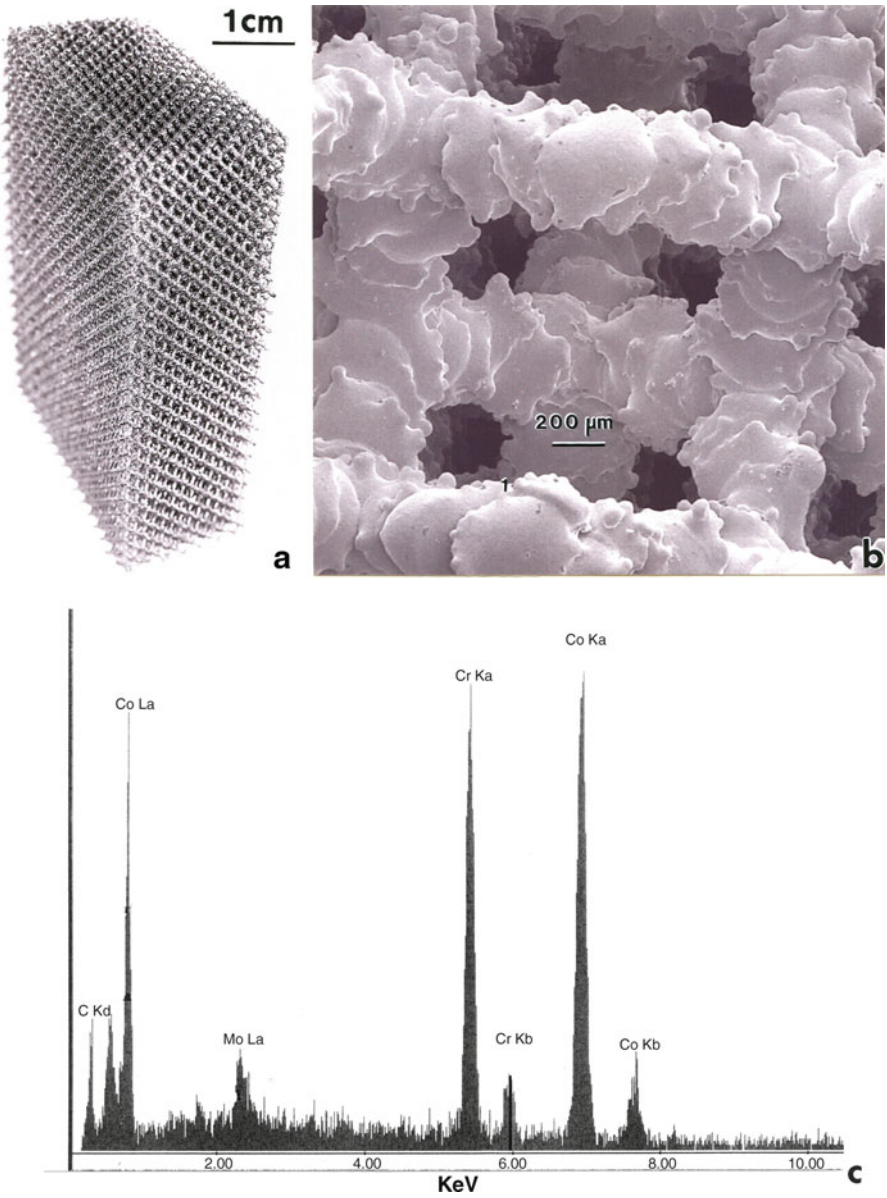


Fig. 21 EBM-fabricated reticulated mesh of Co-26Cr-6Mo. (**a**) Mesh component. (**b**) SEM image of mesh strut structure in (**a**). (**c**) Energy-dispersive X-ray spectrum corresponding to (**b**)

Gibson and Ashby (1982) similarly considered bone structure in particular to consist of short beams or struts (or Timoshenko-type beams (Timoshenko and Gene 1972) similar to those implied in Eq. 2, but in a simple, classical mechanics representation shown in Fig. 25 for a cubic cell similar to the CAD software build unit for fabricating

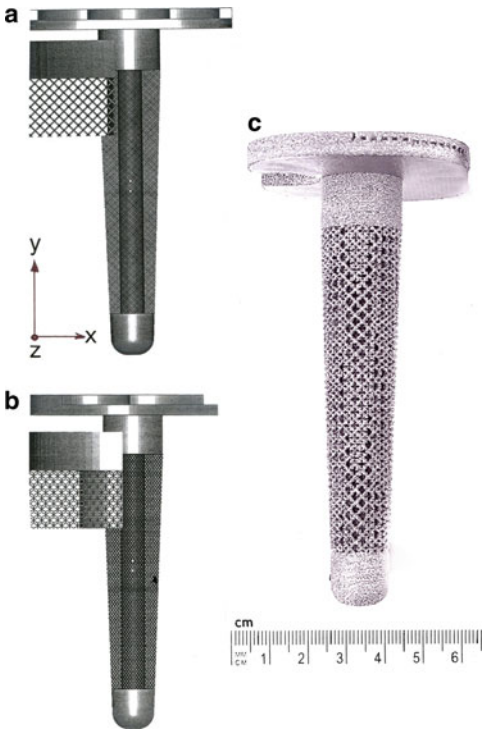


Fig. 22 Tibial (knee) stem designed ((**a**) and (**b**)) and fabricated as a Ti-6Al-4V monolith consisting of an inner rod and surrounded by a reticulated mesh. (**a**) and (**b**) are CAD models rotated 45° (Adapted from Murr et al. (2012b))

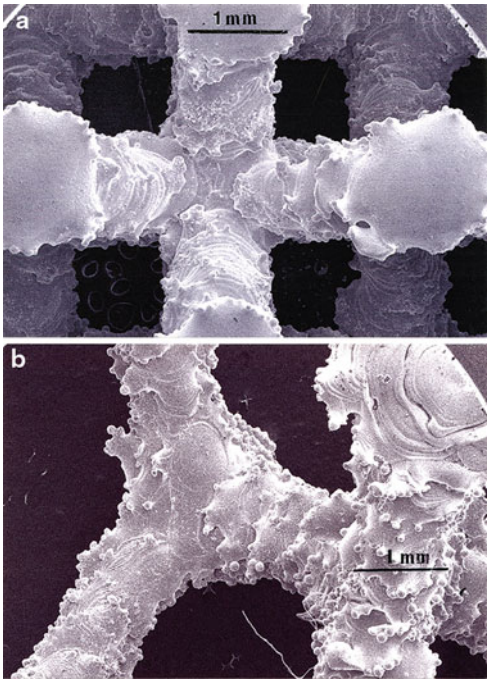


Fig. 23 SEM images for EBM-fabricated Ti-6Al-4V open-cellular structure prototypes. (**a**) Reticulated mesh strut structure. (**b**) Open-cellular foam ligament structure

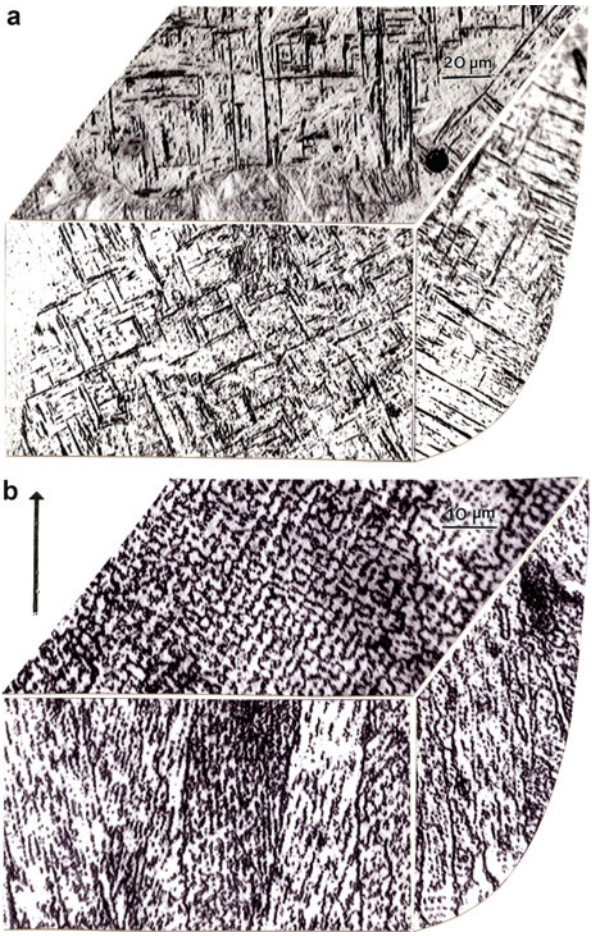


Fig. 24 Reticulated mesh strut 3D section microstructures for EBM-fabricated prototypes. (**a**) α′-martensite in Ti-6Al-4V mesh strut. (**b**) $Cr_{23}C_6$ precipitate columns in Co-26Cr-6Mo-0.2C mesh strut (Adapted from Murr et al. (2011))

the reticulated mesh structures shown in Figs. 21a and 22. In this convention, the second moment of area of a rectangular strut member, I, and the relative density ((ρ/ρ_s) are related to the beam member (or strut) dimensions in Fig. 25 by:

$$\rho/\rho_s \propto (t/\ell)^2 \tag{4}$$

$$I \propto t^4 \tag{5}$$

where t and ℓ are the strut dimensions. The corresponding Young's modulus or stiffness, E, for the cubic, open-cellular (strut) element shown in Fig. 25 is given by Gibson and Ashby (1982):

$$E = CE_sI/t^4 = CE_st^4/\ell^4, \tag{6}$$

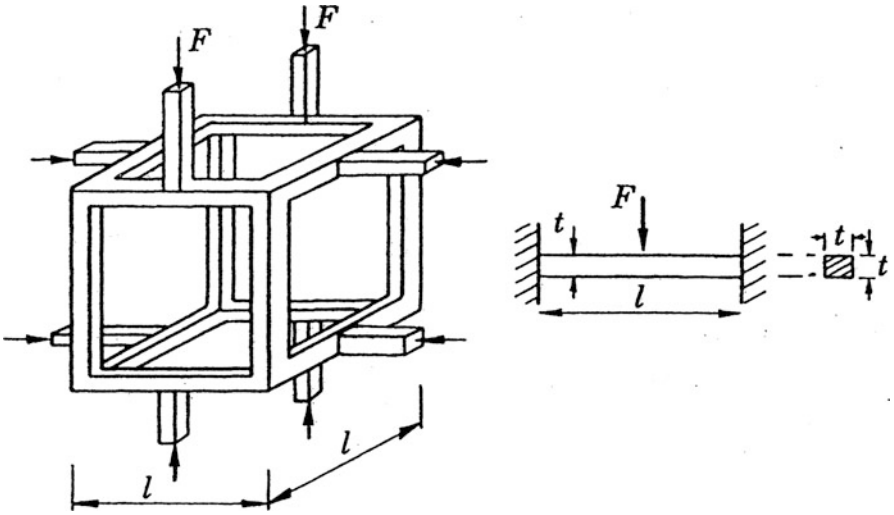


Fig. 25 Simple open-cellular structure element loaded with a compressive force F (Adapted from Gibson and Ashby (1982))

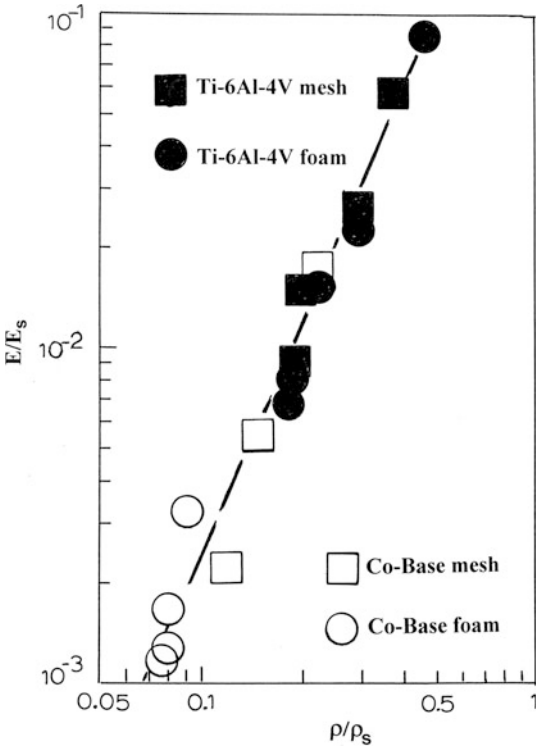


Fig. 26 Relative stiffness versus relative density plots for Ti-6Al-4V and Co-26Cr-6Mo mesh and foam samples fabricated by EBM (After Murr et al. (2011))

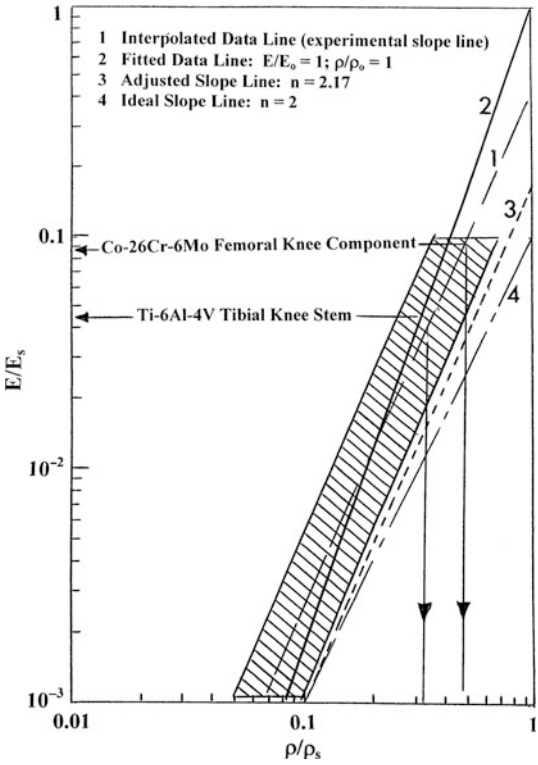


Fig. 27 Stress-shielding design graph. *Shaded region* corresponds to the Ti-6Al-4V and Co-26Cr-6Mo data spread in Fig. 27. Fitted lines are shown by 1–4 in the key box (*upper left*)

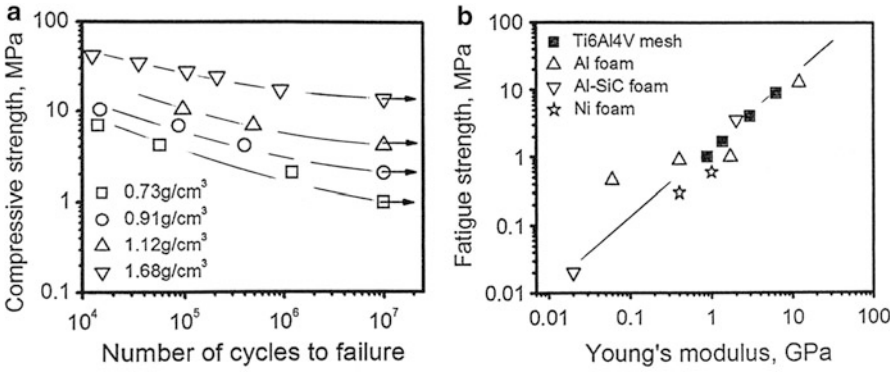


Fig. 28 Fatigue properties for Ti-6Al-4V mesh components fabricated by EBM. (**a**) Compression strength versus fatigue cycles to failure. Compression fatigue rate was 10 Hz. (**b**) Fatigue strength versus Young's modulus, E, for Ti-6Al-4V mesh components compared to foam materials (Adapted from Li, et al. (2012))

Table 2 Nominal mechanical properties of selected biomedical Ti alloys[a]

Alloy	Phase[b] type	Modulus E (GPa)	Yield strength σy (GPa)	UTS (GPa)	Elongation (%)	Hardness Vickers (GPa)
Ti (Grade 4)	α	104	0.49	0.55	15	
Ti6Al-4V	α+β	110	0.84	0.92	8	3.5
Ti-6Al-7Nb	α+β	114	0.91	0.98	12	
Ti-5Al-1.5β	α+β	110	0.88	1.00	16	
Ti-15Zr-4Nb-4Ta-0.2Pd	α+β					
Annealed		89	0.80	0.86	21	
Aged		103	1.00	1.11	10	
Ti-15 Mo (annealed)	β	78	0.54	0.87	21	
Ti-13Nb-13Zr (aged)	β	83	0.87	1.00	13	
Ti-15Mo-5Zr-3Al (aged)	β	69	0.84	0.85	25	
Ti-29Nb-13Ta-4.6Zr (aged)	β	80	0.86	0.91	13	
Ti-35.3Nb-5.1Ta-7.1Zr	β	55	0.55	0.60	19	
Ti-24Nb-4Zr-7.5Sn	β	42	–	0.85	5	2.0
Ti-20Nb-4Zr-7.5Sn	β+α″	75	–	–	–	2.2

[a]Data from Niinomi (1998) and Hao, et al. (2007). Values are averaged (nominal)
[b]α (hcp: a = 3 Å, c = 4.7 Å), β (bcc: a = 3.3 Å), α″ (orthorhombic: a = 3.1 Å, b = 5.1 Å, c = 4.7 Å)

and from Eq. 3

$$E/E_s = C\,(\rho/\rho_s)^2, \tag{7}$$

where the constant, C, can equal 0.9 as in Eqs. 3 and 7 is of the form given previously in ▶ Eq. 1 in chapter "Examples of Natural Composites and Composite Structures".

Ashby et al. (2000) have shown that for a wide range of aluminum and aluminum alloy foams, the exponent in Eqs. 3 and 7 varied from ~1.8 to 2.2, while recent studies of other metal and alloy foams (including Ti-6Al-4V, Cu, and Co-26Cr-6Mo) exhibited an exponent varying between 2 and 2.3 (Murr et al. 2010; Ramirez et al. 2011). However, plotted together with reticulated mesh samples of these metals and alloys and Ni-Cr alloy, the data, in the aggregate, could conveniently be fitted to an exponent of ~2, with E/E_s and ρ/ρ_s plotted on a log-log scale (Murr et al. 2012c).

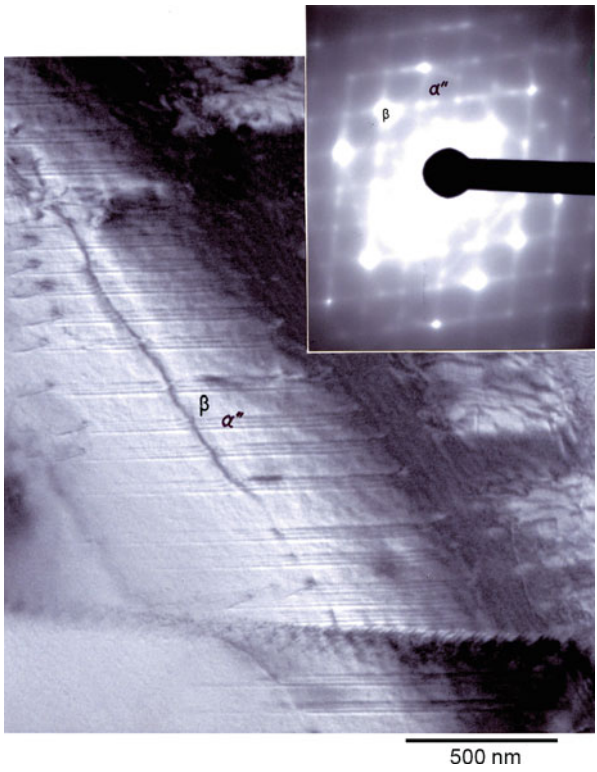


Fig. 29 TEM bright-field image for Ti-24Nb-4Zr-7.9Sn fabricated by EBM, containing α″-martensite plates

Figure 26 shows data of Murr and coworkers (Murr et al. 2011) for Co-26Cr-6Mo and Ti-6Al-4V open-cellular structures fabricated by EBM and fitted to a slope of 27, while Fig. 27 shows a stress-shielding design graph based on this data as shown shaded. The Co-26Cr-6Mo and Ti-6Al-4V lines represent open-cellular (mesh or foam) appliances having an equivalent Young's modulus of 19 GPa. This is based on a solid Young's modulus (or stiffness) of 210 GPa for the Co-26Cr-6Mo alloy and 110 GPa for the Ti-6Al-4V alloy. The corresponding solid densities were 8.44 g/cm^3 and 4.43 g/cm^3, respectively. The data in Fig. 26 were obtained from dynamic modulus (or stiffness) measurements utilizing a nondestructive, resonant frequency or vibration mode established by tapping of an optimum specimen size according to the expression (Li et al. 2012):

$$E = \xi \, m f_r^{\,2} \tag{8}$$

As noted earlier, the fatigue of the fixation polymer (such as PMMA) may also apply to thick mesh or foam structures which have generally lower fatigue strength than solid metals or alloys. Figure 28 shows recent fatigue measurements for Ti-6Al-4V mesh samples corresponding to those illustrated in Fig. 26 (Li et al. 2012). On considering the compression strength in Fig. 28a in contrast

to the fatigue strength versus stiffness (or Young's modulus) in Fig. 28b, it is observed that the implant design strategy may include maximizing fatigue and compressive strength along with compatible stiffness. Consequently, the optimal strategy may involve the use of Ti alloys with lower solid Young's modulus. As observed in Table 2, β-phase Ti alloys, especially more recent Ti-Nb-Zr-Sn quaternary alloys such as Ti-24Nb-4Zr-8Sn (Ti-2448), can have a Young's modulus for the fully solid component as low as 42 GPa. This particular alloy (Ti-2448) has been fabricated in solid monolithic component prototypes by EBM as illustrated in the TEM image of Fig. 29, which shows the formation of α″-martensite (orthorhombic: a = 3.1 Å, b = 5.1 Å, c = 4.7 Å) in the β-phase (bcc: a = 3.3 Å) matrix, where $[100]_{\beta} \parallel [100]_{\alpha''}$ (Hernandez et al. 2013). The α″-phase results because of the rapid cooing in EBM fabrication and produces a microindentation (Vickers) hardness of 2.5 GPa, which is ~25 % harder than conventionally fabricated, β-phase Ti-24 Nb-4Zr-8Sn (Hao et al. 2007).

Although we have mentioned ceramic implants previously (Al_2O_3 and ZrO^2 in particular), Al_2O_3 implants have been shown not to perform satisfactorily as a viable alternative to Ti or Ti alloy implants (Andreiotelli et al. 2009). Zirconia, however, may have potential as an implant material, especially as oral (dental) implants or implant components (Andreiotelli et al. 2009).

References

Andreiotelli M, Wenz HJ, Kohal R-J (2009) Are ceramic implants a viable alternative to titanium implants? A systematic literature review. Clin Oral Implants Res 20(Suppl 4):32–47

Ashby MF, Evans A, Fleck NA, Gibson LJ, Hutchinson JW, Wadley HNG (2000) Metal foams: a design guide. Butterworth – Heinemann, Boston

Baleari M, Cristofolini L, Mineri C, Toni A (2003) Fatigue strength of PMMA bone cement mixed with gentamicin and barium sulfate versus pure PMMA. Proc Inst Mech Eng H217:9–12

Brunski JB, Puleo DA, Nanci A (2000) Biomaterials and biomechanics of oral and maxillofacial implants: current status and future developments. J Oral Maxillofac Implants 15:15–46

Camacho N (2012) Wear characterization of ultra-high molecular weight polyethylene reinforced with multiwall carbon nanotubes. Ph.D. Dissertation, University of Texas at El Paso, El Paso

Egli E (1972) Design properties of structural foam. J Cell Plast 8:245–251

Freyman TM, Yannas IV, Gibson LJ (2001) Cellular materials as porous scaffolds for tissue engineering. Adv Eng Mater 6:369–376

Gibson LJ, Ashby MF (1982) The mechanics of three-dimensional cellular materials. Proc R Soc Lond Math Phys Sci 383:43–49

Hao YL, Li SJ, Sun SY, Zheng CY, Yang R (2007) Elastic deformation behavior of Ti-24 Nb-4Zr-7.9Sn for biomedical applications. Acta Biomater 3:277–286

Hernandez J, Li SJ, Martinez E, Murr LE, Pan XM, Amato KN, Cheng XY, Yang F, Terrazas CA, Gaytan SM, Hao YL, Yang R, Medina F, Wicker RB (2013) Microstructures and hardness properties for β-phase Ti-24 Nb-4Zr-8Sn alloy fabricated by electron beam melting. J Mater Sci Technol 29(11):1011–1017

Karageorgiou V, Kaplan D (2005) Porosity of 3D biomaterial scaffolds and osteogenesis. Biomaterials 26:5474–5491

Li SJ, Murr LE, Cheng XY, Zhang ZB, Hao YL, Yang R, Medina F, Wicker RB (2012) Compression deformation behavior of Ti-6Al-4V alloy with cellular structures fabricated by electron beam melting. J Mech Behav Biomed Mater 16:153–162

Murr LE, Gaytan SM, Medina F, Lopez H, Martinez E, Machado BI, Hernandez DH, Martinez L, Lopez MI, Wicker RB, Bracke J (2010) Next-generation biomedical implants using additive manufacturing of complex, cellular and functional mesh arrays. Philos Trans R Soc A 368:1999–2032

Murr LE, Amato KN, Li SJ, Tian YX, Cheng XY, Gaytan SM, Martinez E, Shindo PW, Medina F, Wicker RB (2011) Microstructure and mechanical properties of open-cellular biomaterials prototypes for total knee replacement implants fabricated by electron beam melting. J Mech Behav Biomed Mater 4:1396–1411

Murr LE, Gaytan SM, Martinez E, Medina F, Wicker RB (2012a) Fabricating functional Ti-alloy biomedical implants by additive manufacturing using electron beam melting. J Biotechnol Biomater 2(3):1–11

Murr LE, Gaytan SM, Martinez E, Medina F, Wicker RB (2012b) Next generation orthopaedic implants by additive manufacturing using electron beam melting. Int J Biomater Article 10245727, 14 p

Murr LE, Gaytan SM, Ramirez DA, Martinez E, Hernandez J, Amato KN, Shindo PW, Medina FR, Wicker RB (2012c) Fabrication of metal and alloy components by additive manufacturing: examples of 3D materials science. J Mater Sci Technol 28(l):1–14

Niinomi M (1998) Mechanical properties of biomedical titanium alloys. Mater Sci Eng A243:231–236

Prawoto Y (2012) Seeing auxetic materials from the mechanics point of view: a structural review of negative Poisson's ratio. Comput Mater Sci 58:140–153

Ramirez DA, Murr LE, Martinez E, Hernandez DH, Martinez JL, Machado BI, Medina F, Wicker RB, Frigola P (2011) Open-cellular copper structures fabricated by additive manufacturing using electron beam melting. Acta Mater 59:4088–4099

Ryan G, Pandit A, Apatsidis DP (2006) Fabrication methods of porous metals for use in orthopaedic applications. Biomaterials 27(13):2651–2670

Timoshenko S, Gene JM (1972) Mechanics of materials. Van Nostrand-Reinhold, New York

Vasilev K, Cooke J, Griesser HJ (2009) Antibacterial surfaces for biomedical devices. Expert Rev Med Devices 6:553–567

Walker PS (1978) Human joints and their artificial replacement. C.C. Thomas, Springfield

Tissue Engineering Scaffolds and Scaffold Materials

Contents

Abstract

This chapter describes concepts of tissue engineering and scaffold fabrication and function as these relate to 3D cell and tissue growth and function which may lead to complex organ manufacture. Scaffold materials are described, including natural (biological) scaffold materials such as collagen, chitosan, and silk as well as metals and polymers.

Introduction

Tissue cells, aside from blood cells, are anchorage dependent, residing in a solid matrix, referred to as an extracellular matrix (ECM). Human tissue, including bone tissue, usually has different ECM types with multiple components and tissue-specific composition. Tissue engineering involves the creation of a scaffold structure having necessary biocompatibility, mechanical properties, especially bone or host tissue matching stiffness, porosity and pore or open-cellular structure, and bioresorbability; or the ability to degrade with time in vivo at a controlled rate in order to leave space for new tissue growth. Scaffolds, cells, and growth-stimulating signals are generally referred to as the tissue-engineering triad. Of course, open-cellular structures associated with permanent implant components as illustrated in ▶ Figs. 7 and ▶ 9 of chapter "Implant Materials and Structures", and especially ▶ Figs. 13, ▶ 14, and ▶ 15 of chapter "Implant Materials and Structures", as well as ▶ Figs. 19 and ▶ 23 of chapter "Implant Materials and Structures" can also provide a tissue-engineering scaffold. Particularly important for prospects for implanting

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_33

open-cellular structures to accommodate the intramedullary, bone volume is the allowance for blood vessel formation (angiogenesis) throughout the scaffold. Current permanent implants, especially intramedullary stems or rods (► Figs. 2, ► 5, and ► 7 of chapter "Implant Materials and Structures"), do not allow for any vascularization.

Scaffolds and Scaffold Materials

Scaffold properties of most importance include, as noted previously, pore size or cell size and shape, the connecting strut or ligament thickness, and surface area and structure. ► Figure 24 (chapter "Implant Materials and Structures") illustrates the strut or ligament features, including surface structure, for EBM-fabricated, open-cellular structures, while Fig. 1 illustrates comparative 3D scaffold systems fabricated from Ti-6Al-4V by EBM, using two different build element (CAD software) models to be discussed in more detail in Part XI. It can be noted in Fig. 1a that the strut thickness and porosity are changing and the stiffness varies from 2.7 to 0.78 GPa (for scaffolds (1)–(3) in Fig. 1a which would in fact correspond to the cancellous (intramedullary) bone regime.

Scaffolds in tissue engineering directed toward tissue replacement, restoration, or the regeneration of detective tissue as in cranio-maxillofacial replacement illustrated in ► Fig. 3b (chapter "Implant Materials and Structures") are often biodegradable polymers such as those illustrated in ► Fig. 2 of chapter "Strategies for Bone Replacement and Tissue Augmentation". However, natural scaffold materials (aside from autografts and allografts of cancellous and cortical bone) include collagen, chitosan (a polysaccharide derived from chitin and found in crab shells) silk, fibrin, and coral. In fact coral has proven to be an effective clinical alternative to autogenic and allogenic bone grafts for some applications (Wong and Mooney 1997). These scaffolds, along with polymer scaffolds, can be created by solid freeform fabrication (SFF) or rapid prototyping (RP) technologies which characterize other forms of 3D printing or additive layer manufacturing to be discussed in detail in Part XI following. In this regard, biodegradable Mg and Mg alloy scaffolds can be fabricated by EBM or a similar build technology using a laser beam to melt precursor (and pre-alloyed) powders in an inert atmosphere, such as argon or nitrogen. This method is described as selective laser melting (SLM) in contrast to EBM fabrication (Part XI, chapter "► Laser and Electron Beam Melting Technologies").

Major advancements in bone tissue engineering in particular have been achieved through the loading of scaffolds or scaffold material with growth factors, drugs, and gene delivery (Bose et al. 2012). Bone replacement scaffolds are similarly loaded with drugs and growth factors to treat bone defects along with introducing osteoinductivity, a property of materials that induces new bone formation through biomolecular signaling and recruitment of progenitor cells. Commonly used drugs include vancomycin, gentamicin, and methotrexate. These tissue enhancement strategies can also be incorporated into hydrogel matrices along with cells which

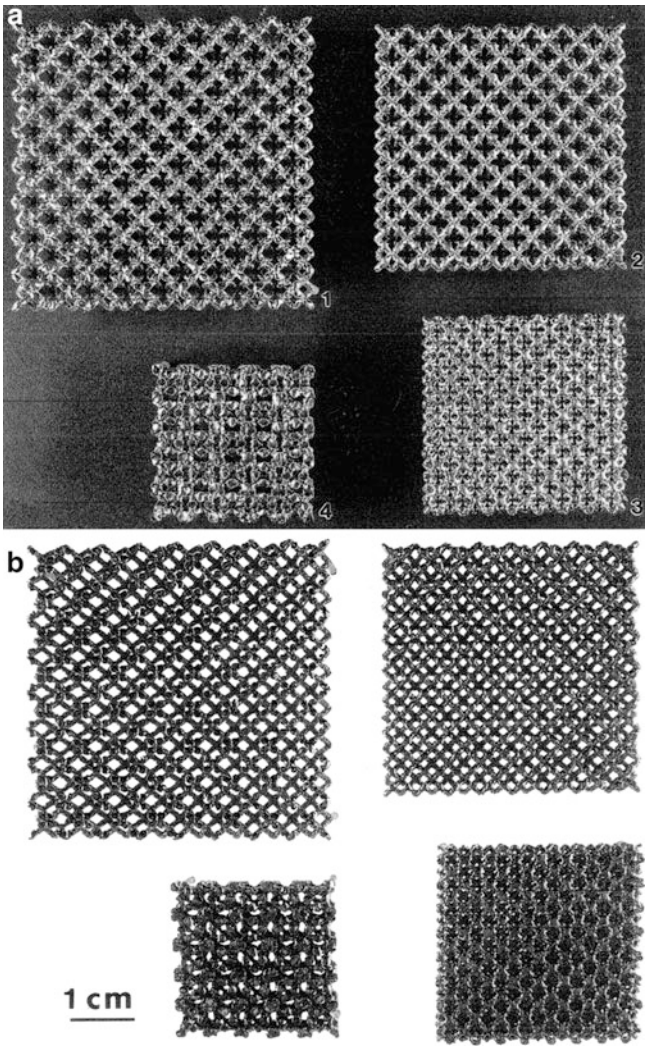


Fig. 1 3D Ti-6Al-4 V scaffolds fabricated by EBM. (**a**) 3D-layer (~3 mm thick) opaque view. *1–3* are same software element as that used to fabricate tibial stem in ▶ Fig. 85c of chapter "Chemical Forces: Nanoparticles". *4* is a different element. (*1*) 82 % porosity, E = 0.92 GPa; (*2*) 81 % porosity, E = 1.53 GPa; (*3*) 72 % porosity, E = 2.70 GPa; (*4*) 65 % porosity, E = 6.78 GPa. (**b**) Transparent view of scaffolds in (**a**) showing pore-channel structure

can be injected into the porous scaffold volume, including vascular cells, since angiogenesis throughout the scaffold (including permanent implant scaffolds) is critical to the success of the scaffold because cells must be within a few hundred microns of the nearest blood supply in order to survive (Karp et al. 2003).

Figure 2 illustrates the cell attachment and tissue development for osteoblast bone cells on a PLA nanofiber mat as a scaffold. The polymer fibers were coated

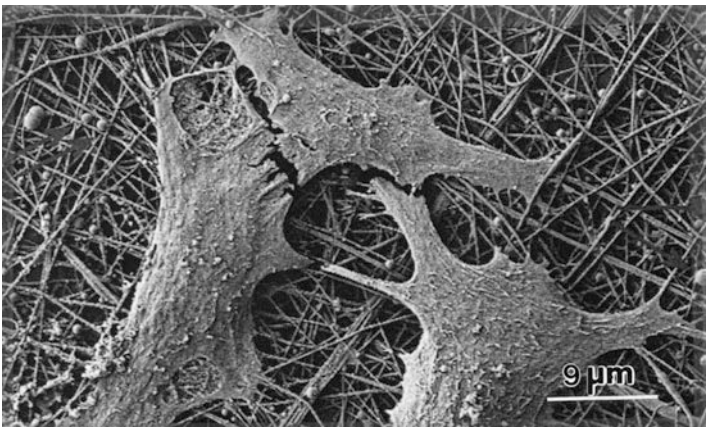


Fig. 2 SEM image of osteoblast bone cells on antimicrobial-coated PLA/PGA (polylactic acid/polyglycolic acid)

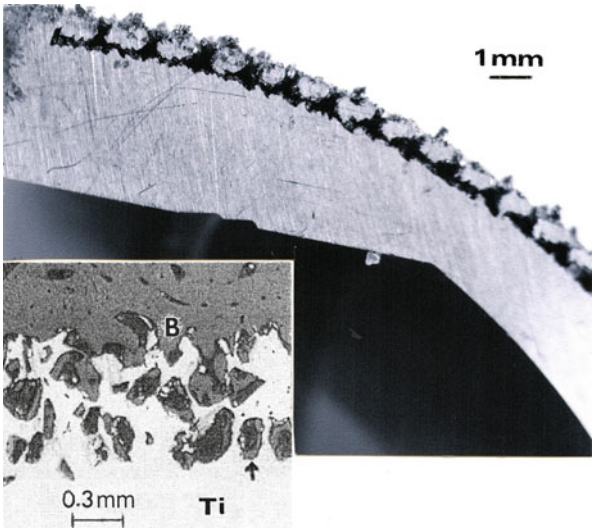


Fig. 3 Section view of Ti-6Al-4V acetabular cup (► Fig. 69 of chapter "Chemical Forces: Nanoparticles") fabricated with a mesh (porous) surface by EBM. Insert shows SEM image of rabbit bone ingrowth (*B*) into a porous Ti implant. *Arrow* shows bone cells in internal pores (Adapted from Reis de Vasconcellos et al. (2010))

with a microgel loaded with an antimicrobial peptide to inhibit bacterial colonization of the scaffold while promoting adhesion, spreading, and proliferation of the osteoblasts. Figure 3 shows a cross-section view of the acetabular cup shown previously in ► Fig. 7 of chapter "Implant Materials and Structures". The insert illustrates how bone cell ingrowth can occur in scaffold systems of this type.

As we noted previously, both tissue engineering and tissue healing have a common concern: the necessity to build new blood vessels throughout the new

or rebuilt tissue. This is particularly true for the rebuilding of skin, the largest body organ. Vascularization is required to keep tissue, such as skin, alive. Skin is composed of two layers: the epidermis (or outer layer) and the dermis (the inner layer). Skin functions to keep vital fluid in the body and harmful pathogens out. It also helps to regulate body temperature. The dermis contains the blood vessels, nerves, sweat glands, oil, and hair. The dermis is composed primarily of connective tissue fibers such as collagen, and it provides nourishment to the epidermis.

Tissue engineering and biotechnology have pursued the creation of suitable and optimal skin substitutes since there are roughly two million people annually who suffer various degrees of burns, frostbites, and disruption of adequate vascularization or venus macrocirculation which cause trophic skin lesions. As early as 500 BC, Hindu surgeons performed nose reconstruction involving grafting skin flaps from the patient's nose. *Xenografts* (skin substitutes harvested from animals for use as temporary grafts in humans) for wound coverage have been reported to have used frog skin as early as 1500 BC, while porcine skin is the most common xerograft in modern practice. *Autografts* (or skin grafts constructed from a patient's own skin) as well as *allografts* (skin grafts constructed from cadaver skin) were the major forms of skin repair until the late twentieth century.

The first synthetic skin was invented by John Burke and Ioannis Yannas in the 1970s who used collagen fibers and a long sugar molecule to form a porous polymer material resembling skin (Yannas and Burke 1980). They later created an artificial skin using polymers from shark cartilage and collagen from cowhide (Sheridan et al. 1994). This created a membrane covering similar to the human dermis layer, and a layer of silicone was added to act like the human epidermis. This synthetic skin functioned like a scaffold onto which new skin and blood vessels could grow, and as this process proceeded, the scaffold degraded and was absorbed by the body.

In the mid-1980s, Joseph Vacanti and Robert Langer conceived the idea of constructing a biodegradable scaffold to grow skin cells using fibroblast cells extracted from neonatal foreskin removed during circumcision (Longer and Vacanti 1995). One foreskin can yield enough cells to make roughly four acres of grafting material. The fibroblasts can be grown on biodegradable scaffolds as illustrated in Fig. 2.

A collagen matrix or mixtures of collagen and fibrin can also be used as even more biocompatible scaffolds, and cells can also be incorporated in the collagen matrix. Similarly, customized gels and hydrogels have been utilized to regrow skin with vascularization. These can be fabricated as tissue coverings or as growth matrices and can include collagen-based gels or hydrogels composed of dissolved dextran and polyethylene glycol (PEG) (Sun et al. 2011). Such hydrogels can act as scaffolds for other soft tissue growth and propagation such as nerve cells and the like.

While skin can be regenerated and grown on tissue-engineering scaffolds, it represents a shallow (or thin) 3D organ in contrast to other more complex organs which are composed of multiple cell types, extracellular matrices of proteins, and special microstructures (including tubular and sheet forms) which exhibit

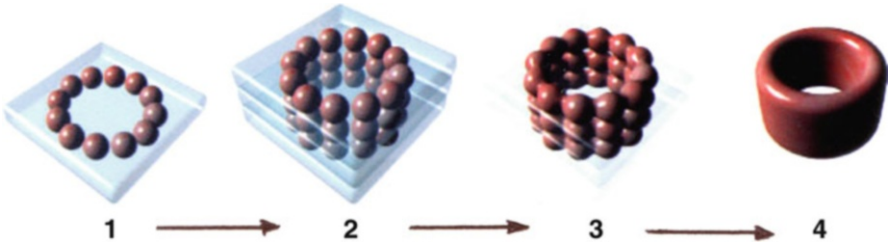


Fig. 4 3D bioprinting of organ or tissue structure using bioink spheroids of cells or aggregated multicell types injected into water-based biopaper. (*1*) bioink spheroids printed into a layer of biopaper gel. (*2*) additional layers printed to build the object. (*3*) bioink spheroids fuse together; biopaper dissolves. (*4*) final living tissue product which can include speciated cellular regimes forming functional components formed by self-assembly (Adapted from an Organovo, Inc. graphic (Courtesy of Organovo, Inc.))

structure-specific biological functions. In order to fabricate these larger, more complex organs, or in fact to fabricate skin in a manufacturing process such as biofabrication or bioprinting, Makoto Nakamura, around 2002, conceived the idea of incorporating living cells in a spherical hydrogel emulating an inkjet printer ink droplet. By 2008, Nakamura and co-workers had created a working bioprinter that could print biotubing similar to a blood vessel (Nakamura et al. 2005). About the same time, Gabor Forgacs at the University of Missouri, along with his co-workers, established a commercial bioprinting company, Organovo, which printed functional blood vessels and cardiac tissue using cells from a chicken Mironov et al. 2007, 2009). In bioprinting processes, aggregates of specialized living cells create bioink spheroids which are injected by a printer head onto a substrate or biopaper made from collagen, gelatin, or other hydrogels. As more layers of bioink spheroids are selectively added layer-by-layer, the water-base substrate (biopaper) dissolves, and the bioink cell spheroids fuse together as illustrated schematically in Fig. 4.

It has been shown that specific cells contained in bioink spheroids are capable of rearranging themselves (self-assembly) after printing. As an example, experimental blood vessels have been bioprinted using such bioink spheroids comprised of aggregates of fibroblast cells, smooth muscle cells, and endothelial cells. Once printed as illustrated in Fig. 4, the endothelial cells migrate to the inside of the bioprinted blood vessel, the smooth muscle cells move to the middle of the wall, while the fibroblasts populate the outer wall (Forgacs 2012).

The implication of this technology is that human organs might be fabricated by 3D printing or additive manufacturing. Correspondingly, organs specific to an individual might be manufactured using micro-CT scans to embed the organ morphology in CAD software, while either using specific donor cells or the speciation of stem cells with donor DNA to produce a patient-specific organ similar to fabricating patient-specific implant components as described previously. We will discuss these prospects in more detail in Part XI.

References

Bose S, Roy M, Bandyopadhyay A (2012) Recent advances in bone tissue engineering scaffolds. Trends Biotechnol 30(10):546–554

Forgacs G (2012) Tissue engineering: perfusable vascular networks. Nat Mater 11:746–747

Karp JM, Dalton PD, Shoichet MS (2003) Scaffolds for tissue engineering. MRS Bull 28:301–306

Longer R, Vacanti JP (1995) Tissue engineering. Science 14:920–921

Mironov V, Prestwich G, Forgacs G (2007) Bioprinting living structures. J Mater Chem 17:2054–2060

Mironov V, Visconti RP, Kasyanov V, Forgacs G, Drake CJ, Markwald RR (2009) Organ printing: tissue spheroids as building blocks. Biomaterials 30:2164–2174

Nakamura M, Kobayashi A, Takagi F, Watanabe A, Hiruma Y, Ohuchi K, Iwasaki Y, Hovie M, Morita I, Takatoni S (2005) Biocompatible inkjet printing technique for designed seeding of individual living cells. Tissue Eng 11:1658–1666

Reis de Vasconcellos LM, Leite DO, de Oliveira FN, Carvalho YR, Alves Cairo CA (2010) Evaluation of bone ingrowth into porous titanium implant: histomorphometric analysis in rabbits. Braz Oral Res 24(4):1–9

Sheridan RL, Hegarty M, Tonpkins RG, Burke JF (1994) Artificial skin in massive burns – results to 10 years. Eur J Plast Surg 17(2):91–93

Sun G, Zhang X, Shen Y, Sebastian R, Dickinson LE, Fox-Talbot K, Reinblatt M, Steenbergen C, Harman JW, Gerecht S (2011) Dextran hydrogel scaffolds enhance angiogenic responses and promote complete skin regeneration during burn wound healing. Proc Natl Acad Sci U S A 108(52):20976–20981

Wong WH, Mooney DJ (1997) Synthesis and properties of biodegradable polymers used as synthetic matrices for tissue engineering. In: Atala A, Mooney DJ (eds) Synthetic biodegradable polymer scaffolds. Birkhausen, Boston, pp 83–95

Yannas IV, Burke JF (1980) Design of an artificial skin. I. Basic design principles. J Biomed Mater Res 14:65–81

Part XI

Advanced Materials Processing and Fabrication: 3D Printing and Additive Manufacturing Technologies

Photolithography Applied to Integrated Circuit (IC) Microfabrication

Contents

Abstract

Photolithography as it relates to lithography fundamentals is described and applied to the lithography process utilized in producing integrated circuits (ICs) and layered electronic device structures. These structures have some thickness or layering restrictions but are nonetheless layer-manufactured structures on an electronic landscape, usually a silicon wafer substrate.

Introduction: Photolithography Fundamentals

Additive manufacturing or 3D printing technologies employ a layer-by-layer building process to fabricate 3D, monolithic structures using computer-aided design (CAD) strategies or models. These technologies are also known as optical fabrication, photo-solidification, or solid freeform fabrication (SFF). More generally these technologies are referred to as *stereolithography* (SLA), a term coined by Charles Hull in 1986. However, stereolithography can be conceptually related to a lineage of lithographic processes.

Lithography, a method for printing using limestone (or other flat stones; from the Greek lithos, meaning stone, and graphein, meaning to write), was invented by Alois Senefelder around 1797 and used by artists to produce prints of their work. In this process, the artist drew a picture on the surface of a flat or flattened stone (primarily limestone) with a mixture of wax, soap, lampblack (the ink), and rainwater. This image would repel water, while the surrounding, uncoated stone surface would retain the water, selective hydrophobic/hydrophilic surface regimes.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_34

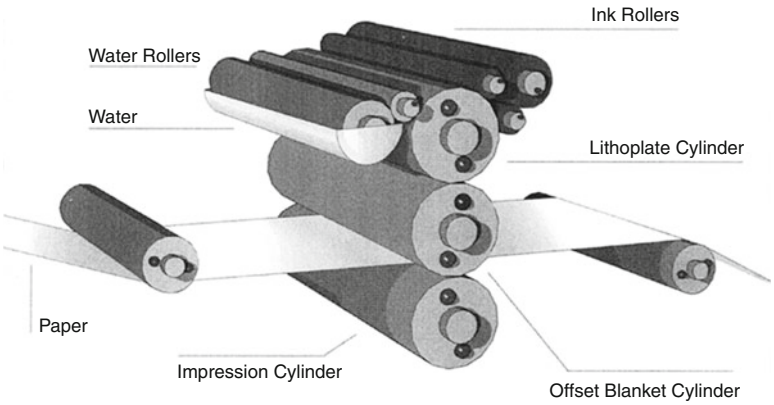


Fig. 1 Offset lithography schematic view

The hydrophobic (image) regions could be recoated with ink in the greasy (wax) matrix to allow continuous printing or transfer of the inked image to paper or other suitable medium. However, the image in normal lithography is reversed, but could be offset by transferring the image to a flexible sheet such as rubber and then printed from the offset (reversed) image. This is the standard printing format today except the image is made of a polymer coating applied to a flexible aluminum plate affixed to a roller-plate system as illustrated in Fig. 1. This modern offset lithography or high-volume lithography used to produce newspapers, magazines, books, posters, and other paper image formats depends on photography or a photographic process. In the original photolithography process, images were photographically transferred to a stone surface or the more efficient, flexible aluminum plate. This process led to breaking down the original photograph into dots of varying sizes suitable to press reproduction which became known as the halftone process for reproduction.

Color printing of multicolor prints or chromolithography used separate stones or aluminum plates for each color, and the paper or medium to be printed upon must pass through the printing press for each color in the final print. Precise registry is required to assure that each color is placed in the precise position in each print. These are also broken down in the halftone process producing printable color dots, with image resolution determined by the number of dots printed in a unit area: usually dots per inch (dpi) in an area one inch on a side. In the preparation of the lithographic plates, the polymer coating is photosensitive and resistant to acids or other etching compounds. This material is referred to as a photoresist, which when exposed to the photographic negative hardens where the negative image allows light to pass. The photoresist is developed by rinsing in a solvent which removes the unhardened areas. The image area on the prepared plate takes up ink as noted previously. In offset printing (or lithography) as illustrated in Fig. 1, each color of ink is applied separately – one for each plate. However, color dots of four inks – cyan (blue), magenta, yellow, and black – are deposited in specific patterns, creating color addition which produces a wide range of perceived colors. This is

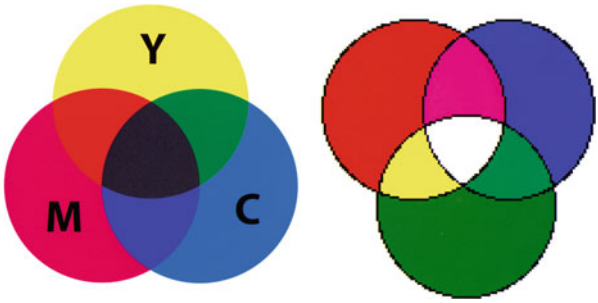


Fig. 2 Color addition. (a) Standard offset *cyan* (*C*), *magenta* (*M*), and *yellow* (*Y*). *Red–green–blue* addition characteristic of display color

referred to as four-color printing or four-color process lithography. Figure 2 illustrates these color fundamentals, which in various ways can be applied to color image formation or projection, including three-phosphor dot image screens or flat screen arrays involving LCD or LED color dots (or areas). Correspondingly, inkjet printing can print color dots in similar patterns to create full color, or the perception of a wide range of color hues when more than four colors are printed. It will be observed later that inkjet arrays can be used to print in three-dimensions as illustrated conceptually in ▶ Fig. 96 of chapter "Chemical Forces: Nanoparticles" for 3D-cell printing or bioprinting.

Microfabrication Fundamentals

Integrated circuits, monolithic integrated circuits (ICs), or what are commonly referred to as microchips are fabricated by multiple, overlapping, functional, patterned layers defined by photolithography. Patterns are created on flat semiconductor surfaces, usually silicon similar to the photoresist sensitization in offset lithography, although color (image) sequences represent electronic functions or functional regions of conductors, insulators, or doping sequences to create memory devices. These are often complementary metal–oxide–semiconductor (CMOS) field-effect transistors which are formed whenever the gate layer (polysilicon or metal) crosses a diffusion layer creating n- or p-type semiconduction region as illustrated schematically in Fig. 3 (Baker 2010).

This process begins by spin coating a silicon wafer with a photoresist similar to the offset lithography plate. A photomask characterizing the circuitry or circuit elements in a small feature image is projected onto the wafer photoresist using intense light, usually ultraviolet (UV) light. In modern IC fabrication, deep UV excimer lasers (krypton fluoride (KrF)) having a wavelength of 248 nm or argon fluoride (ArF) at a wavelength of 193 nm are used. The light source wavelength (λ) is a controlling factor in the minimum feature size that the projection system can print:

$$\mathrm{CD} = \mathrm{K}\lambda/\mathrm{NA}, \tag{1}$$

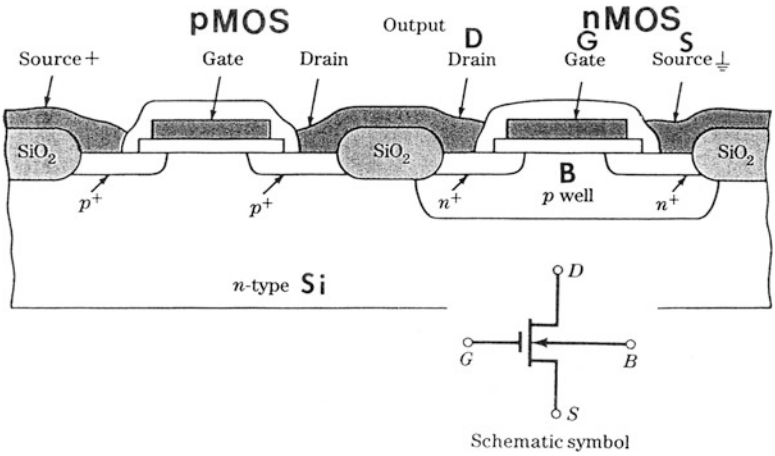


Fig. 3 CMOS (n- and p-type) field-effect transistors fabricated on the same silicon wafer substrate

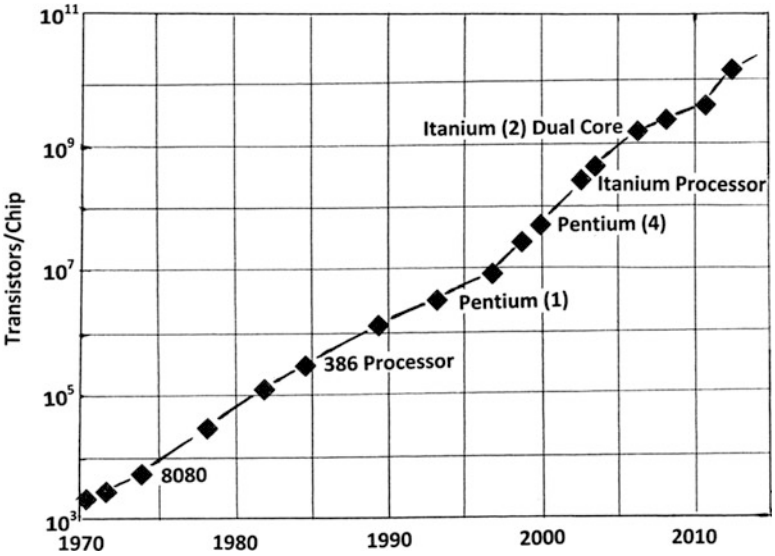


Fig. 4 Moore's law: transistors per chip by production year (Intel data). Only a few types of processor chips are noted

where CD is the critical dimension or minimum feature size, λ is the wavelength, K is a process factor commonly assumed to have a value of ~0.4, and NA is the numerical aperture of the lens relative to the wafer surface. While this characterizes a design role for lateral dimensions of features in the plane of the IC (or the silicon wafer), the depth of focus (DF) presents a competing constraint since it restricts the thickness of the photoresist and the wafer topography or functional layer thickness:

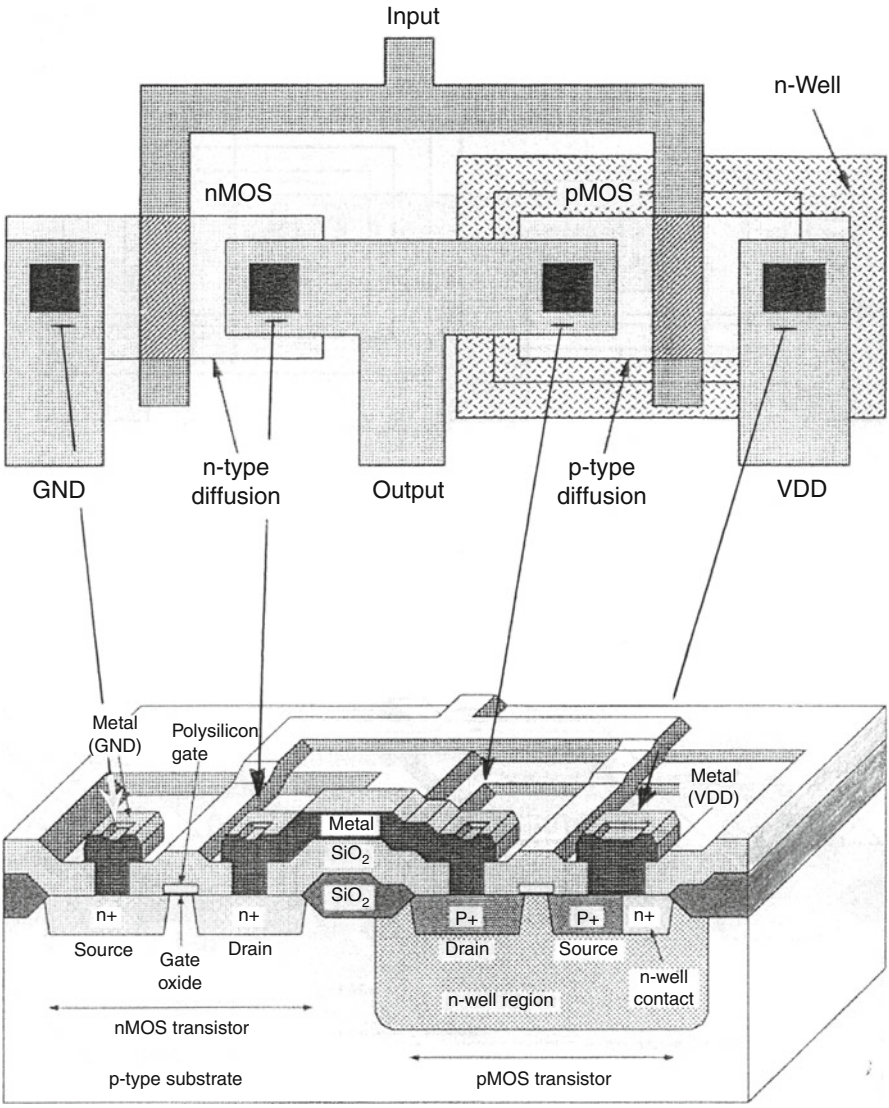


Fig. 5 Cross-sectional view showing CMOS transistor on chip composite layout. GND and VDD denote ground and above ground voltage contacts (Adapted from Maly 1987)

$$\mathrm{DF} = \mathrm{K}^{'}\lambda/(\mathrm{NA})^{2}, \tag{2}$$

where $\mathrm{K}^{'}$ is another process-related parameter.

Generally, CD in ▶ Eq. 1 of chapter "3D Printing: Printed Electronics" is connected with circuit size or the number of transistors (or memory elements) on an IC chip, which has followed the doubling law or Moore's law. In 1965, Gordon Moore, cofounder of Intel, noted that the number of transistors on an IC

chip would double every 18–24 months. This law has been generally followed up to the present time (2013), where the Intel chip density exceeds 10 billion transistors as illustrated in Fig. 4. Correspondingly, minimum feature sizes in chip manufacturing (▶ Eq. 1 of chapter "3D Printing: Printed Electronics") have decreased from 500 nm in 1990 to below 45 nm in 2010, with CD (▶ Eq. 1 of chapter "3D Printing: Printed Electronics") approaching ~10 nm by 2020. The ultimate limit of downsizing will be roughly the distance between silicon substrate atoms, ~0.3 nm. Operating frequency (or clock speed) trends for Intel processors have also increased from ~50 MHz in 1990 to ~3 GHz in 2010. IC circuit size classification also changed from small-scale integration (SSI) in 1963 to medium-scale integration in 1970 (Fig. 4), large-scale integration (LSI) in 1975, very-large-scale integration (VLSI) in 1980, ultra-large-scale integration (ULSI) in 1990, and to giga-scale integration (GSI) after 2010 (Fig. 4). The VLSI revolution began in large part as epitomized by the classic text by Carver Meade and Lynn Conway, "Introduction to VLSI Systems" published in 1980 (Meade and Conway 1980).

Figure 5 illustrates an example of a cross-sectional view of a fabricated chip with a reversed, p-type Si substrate, in contrast to Fig. 9.3, showing an nMOS and pMOS transistor. A modern CMOS wafer chip will go through the photolithographic cycle as many as 50 times. While this represents a feature of layer-by-layer fabrication, the integrated circuit has come to be known as a monolithic integrated circuit or single, functional chip. A truly 3D lithographed electronic monolith could be constructed by stacking chips and interconnecting them by vertical, conducting vias. This is a difficult feature to achieve in a unitized production process.

Photolithography (or optical lithography) as illustrated in the context of UV excimer laser lithography implicit in ▶ Eqs. 1 and ▶ 2 of chapter "Tissue Engineering Scaffolds and Scaffold Materials" and as applied to Fig. 5 shares some fundamental principles with offset color lithography printing (Fig. 1) in that a photographic (mask) pattern is created in a photoresist or etching resist through UV excimer laser exposures with the design feature of each layer function characterized by a color: green, n-diffusion; red, polysilicon; blue, metallizational; yellow/brown, p-diffusion; and magenta (purple) and cyan (light blue), metallizations 2 and 3, respectively. However, this exposure can be accomplished by direct writing of the pattern by the laser or by an electron beam, and the subsequent stages in the lithography process have more in common with etching than lithographic printing (Fig. 1). The photoresist can be positive (preferred) or negative relative to the solubility in the developer of exposed areas.

References

Baker RJ (2010) CMOS: circuit design, layout, and simulation, 3rd edn. Wiley/IEEE, New York

Maly W (1987) Atlas of IC technologies: an introduction to VLSI processes. Benjamin/Cummings, Menlo Park

Meade CA, Conway L (1980) Introduction to VLSI systems. Addison-Wesley, Boston

3D Printing: Printed Electronics

Contents

Abstract

Aside from photolithography to produce electronic integrated circuits and device system, 3D printing can allow these circuits to be directly printed using novel ink compositions or nanoparticles and integrated printer heads. Printed electronics in its various forms and innovations can revolutionize electronics manufacturing. This chapter outlines contemporary developments and future innovations in 3D electronics printing.

Introduction to Printed Electronics

In the case of traditional printed circuit boards, either copper-coated laminates are selectively etched in patterned areas not protected by a resist or the copper conductive pathways (or wiring) are electroplated directly onto the laminate surface. These processes can involve screen printing, offset lithography, gravure printing on flat surfaces, and gravure and flexography printing, and inkjet printing on flexible surfaces, including polymer sheets and paper. Printed electronics has evolved into a new class of materials systems with applications in photovoltaics (solar cell platforms such as flexible solar collector sheets), antenna arrays on curved or flexible surfaces, visual effects and displays, and even disposable electronics/devices. These printed or printable systems can involve inorganic electronics characterized by highly ordered layers and interfaces or organic (polymer) materials, including A.C. electroluminescent (EL) multicolor displays involving 6–8 printed inorganic layers, including copper-doped phosphor, on polymer film substrates. Similarly,

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_35

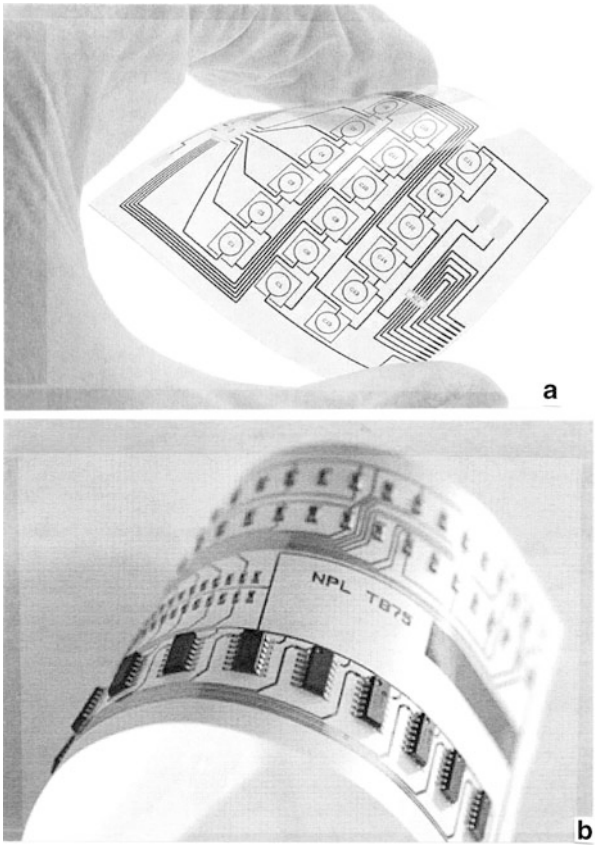


Fig. 1 Printed electronic circuits on thin, flexible, polymer substrates. (**a**) Patterned, inkjet circuits on polymer sheet. (**b**) Surface-mount devices on flexible, patterned electronic circuit (flex circuit) photo lithographically printed

LED inorganic (EL) devices can be similarly printed along with organic field-effect transistors (OFETs) and organic light-emitting diodes (OLEDs) and organic photovoltaic cells (OPVCs). These can be conveniently printed by inkjets along with conducting ink interconnects. Novel configurations can also be printed on fabric landscapes which can incorporate conducting polymer threads as interconnects. Figure 1 shows some examples of printed electronics on flexible substrates to which chips or other surface-mount devices can be attached as well.

One of the more novel aspects of 3D printed electronics involves LED arrays. These arrays can involve inorganic LEDs, organic LEDs, and hybrids fabricated partly from inorganic materials and partly from organic/polymeric materials. Inorganic LEDs based on GaN grown on sapphire substrates (hcp-Al_2O_3) as illustrated schematically in Fig. 2a, or AlGaInP/GaAs LED structures fabricated on GaAs substrates as illustrated in Fig. 2b, require high process temperatures, but are characteristically highly ordered layers and interfaces. They provide high brightness, durability, and energy efficiency (>50 %) and are moisture resistant. Correspondingly, these device structures are difficult if not impossible to fabricate directly on flexible substrates such as polymers or paper, but there are novel

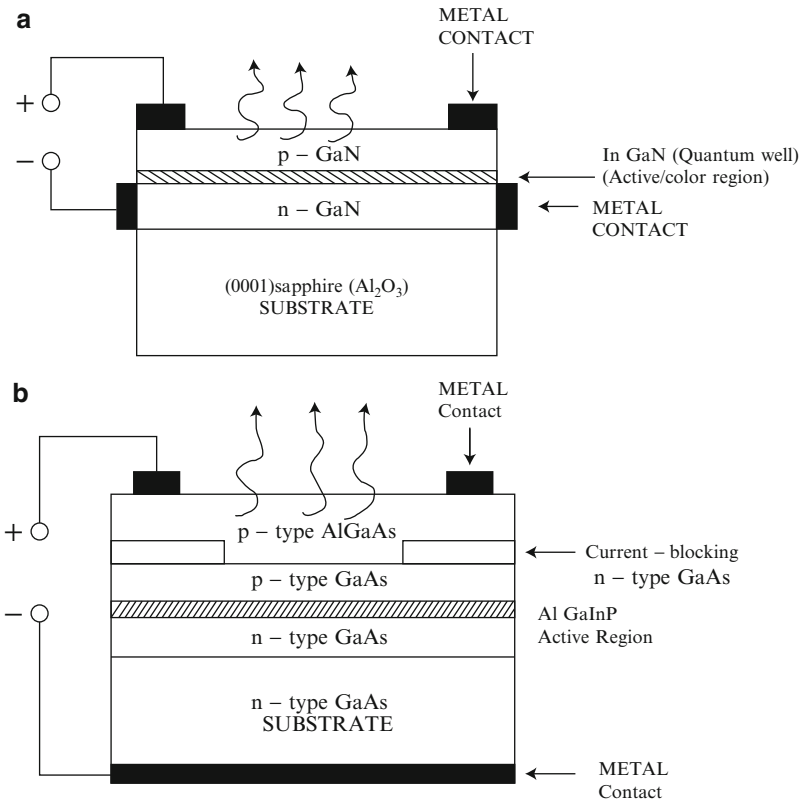


Fig. 2 Conventional, planar GaN/InGaN (420 nm) LED structure (**a**) and AlGaInP/GaAs (650 nm) LED structure (**b**). In (**a**) the active, quantum-well (QW) regime can be composed of n- or p-type ternary regimes with P-GaN interlayers which can produce polychromatic light. In (**b**), the active region can also be structured as a polychromatic p-n sandwich region. Metal contacts are often Ni/Au alloy, Al, Au, or Ti/Ni/Au

concepts to transfer inorganic LEDs to such flexible substrates by, for example, laser liftoff and contact printing (Chun et al. 2012).

While inorganic LEDs can be made as small as ~10 μm on a side, the human eye cannot resolve features smaller than about 100 μm, and sizes smaller than ~100 μm do not improve display quality. In addition, although the human eye is sensitive to light in the visible spectrum, ranging from blue/violet (~400 nm) through red (~700 nm wavelength), the eye is most sensitive to green (~555 nm wavelength) light. Green light also affects the perception of white light (▶ Fig. 2 of chapter "Photolithography Applied to Integrated Circuit (IC) Microfabrication") along with red and blue which are readily available in LEDs, while green is not as well perfected. $In_xGa_{1-x}N$, alloy-based LEDs are the most promising green devices. ZnO, which has the same crystal structure (hexagonal wurtzite) as GaN, can be deposited in place of the GaN layer as illustrated in the white-light generation

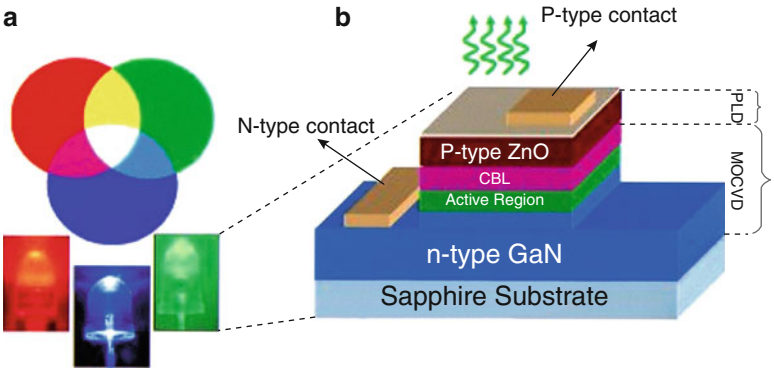


Fig. 3 LED white-light generation by color addition (► Fig. 2 of chapter "Photolithography Applied to Integrated Circuit (IC) Microfabrication") (**a**, **b**) shows a hybrid green LED structure schematic on a sapphire substrate. CBL denotes the AlGaInN current-blocking layer. The underlying active layer is InGaN

concept illustrated in Fig. 3. In this device fabrication, metalorganic chemical vapor deposition (MOCVD) is used to grow the GaN, and InGaN (active) layer as well as the AlGaInN current-blocking layer (indicated by CBL in Fig. 3), while pulsed laser deposition (PLD) growth forms the ZnO layer. These are sequential processes similar to those employed in VLSI and ULSI chip fabrication, but involving only a few steps in contrast to several dozen in chip fabrication.

In contrast to inorganic LED fabrication on flat, crystalline substrates, organic LED fabrication can be accomplished on flexible substrates, including polymer sheets and paper. The flexible polymer substrates can be clear. Organic LED (OLED) structures can be fabricated as single or multilayer (sandwich) devices as illustrated schematically in Fig. 4. The polymer layers can be applied by vapor deposition in vacuum or by low-pressure organic vapor-phase deposition (OVPD). More importantly, inkjet deposition is also a viable route which reduces cost and allows printing of OLED arrays onto very large, flexible substrates. Multilayer or stacked OLED arrays can produce white light as implicit in ► Fig. 2 of chapter "Photolithography Applied to Integrated Circuit (IC) Microfabrication" and Fig. 3.

Typically, OLEDs are ~0.1–0.5 μm thick. In the 2-layer structure shown in Fig. 4b, the conducting layer beneath the anode transports holes from it. As noted, polyaniline (PANI) is a common conducting polymer in this layer. The emissive layer (Fig. 4b) is composed of a different polymeric component (polyfluorene (PFO)) that transports electrons from the transparent cathode (which can be indium tin oxide (ITO) or a conductive, transparent polymer such as PEDOT: PSS) to form the emitted light. OLEDs are often susceptible to oxidation and are usually sealed with a clear polymers encapsulating coating. In addition, the emission efficiency is generally lower than inorganic LED emission, although large-area, flexible, white OLEDs are being manufactured commercially with energy efficiencies 2–3 times higher than incandescent bulbs, and brightness comparable to fluorescent bulbs. While power consumption is at best nearly two orders of magnitude smaller than

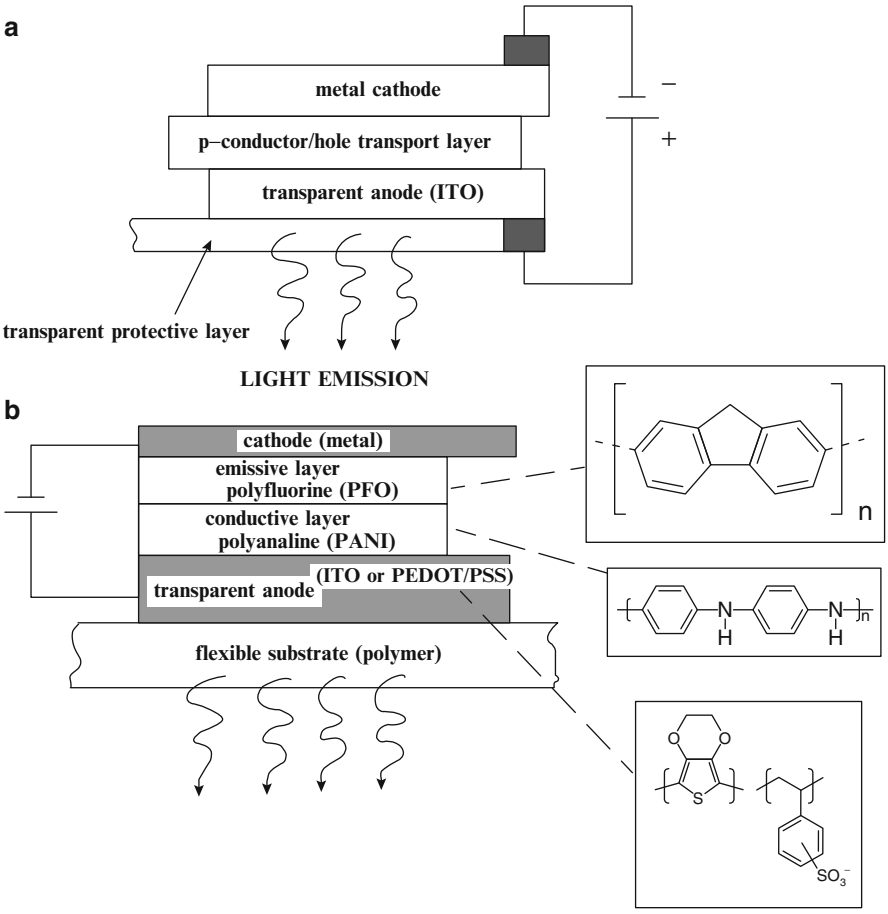


Fig. 4 OLED structure. (**a**) Single-layer (**b**) 2-layer structure illustrating polymer components. PEDOT, PSS represents poly (3, 4-ethylenedioxythiophene): poly(styrenesulfonate), a transparent, conductive polymer, ITO is indium tin oxide, a transparent semiconductor. Polymer structure (monomer) inserts are shown

incandescent lighting to produce the same brightness (~1 W/25 lm output), there is still heat loss which must be removed in order to maintain efficiency of operation. Consequently, thermal management is an important feature in LED design and fabrication strategies.

Since organic, polymeric materials can act as both n- and p-type semiconductors in p-n heterojunction fabrication, it is possible to fabricate n-p-n- and p-n-p-type transistors or organic field-effect transistors (OFETs) which can also act as a unipolar or ambipolar light-emitting OFET (Klauk 2010). Consequently, flexible integrated circuits, particularly organic or polymer-based ICs, can be fabricated, but not with the small feature sizes which are achievable in conventional IC fabrication as described previously.

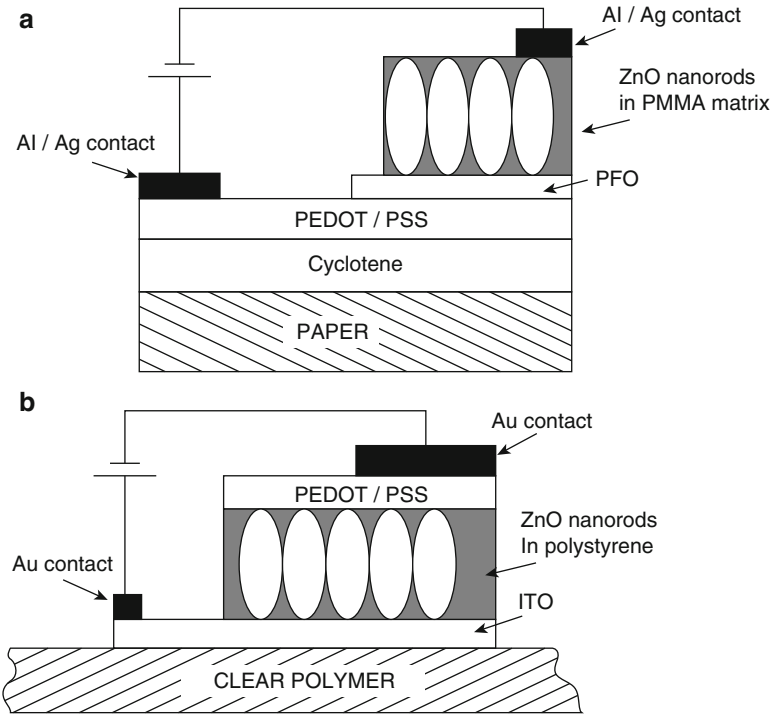


Fig. 5 Hybrid, white-light LED structures. ZnO layer is composed of oriented nanorods in a PMMA matrix (**a**). ZnO in polystyrene (**b**)

The ability to selectively print both inorganic and organic component device functions, especially utilizing multiple inkjet printing heads, will allow large-area flexible substrates, even paper, to be used as new classes of lighting and display systems as well as disposable devices. Novel, hybrid LED structures (Fig. 5) have been printed on paper recently (Repo et al. 2013) which has been coated with a polymeric resin, cyclotene ($C_6H_8O_2$), a dielectric used in commercial microelectronic applications as an insulator. Paper can be problematic for electronics or microelectronics unless either the paper finish or the ink matrix, or both, can be appropriately adjusted to maintain optimal functionality. The ability of ink to partition device functionality – metal conductors, semiconducting regimes, insulating layer application, etc. – will be described in more detail below.

The printing of white, luminous wallpaper for home or office lighting using LED arrays is a particularly intriguing prospect. In addition, it has recently been demonstrated that a printed GaAsGe solar cell can exceed 40 % conversion efficiency. This is roughly eight times that of the most efficient polymeric cells, and nearly the same as more conventional Si-based cells (Repo et al. 2013), and visibly transparent photovoltaics which can be incorporated into window glass, making windows solar collectors (Lunt 2012).

Recent research has also been directed at the development of paper-based polymer batteries using cellulose/polypyrrole (ppy) composite layers sandwiching soft paper (or cellulose) soaked with electrolyte, with Pt metal contacts (Nyström et al. 2009). Polypyrroles are pentagonal ring-type conducting polymers which, along with related conducting polymeric materials, were discovered by A.G. MacDiarmid who received the Nobel Prize in chemistry in 2000 (MacDiarmid 2001). These components can be printed in appropriate ink matrices from selected inkjet heads. As shown later, it is possible to integrate multiple device printing into fabrics or fabric printing as well as a host of other flexible substrates.

Inkjet Printing: Ink Compositions and Properties

In 1867, Lord Kelvin developed the first printing device that used drops of ink emitted from an orifice or nozzle to record modulated lines representing the dots and dashes of Morse code in telegraph transmissions. Hewlett Packard introduced the first desktop inkjet printer for computer hardcopy in 1984. Continuous inkjet printers were the first to be used for printing images from a constant stream of charged ink drops which could be deflected into a recycling system or allowed to impact the paper. The more popular drop-on-demand printers eject single drops of ink as required. A thermal inkjet printer, or bubble jet printer, utilizes a heating element inside the print (inkjet) head to vaporize some of the ink and create a bubble. The bubble forces a drop of ink out of the printer orifice. In drop-on-demand printers, the printer head reservoir may also be heated to create requisite drop surface tension and viscosity, while a piezoelectric device creates pressure pulses to eject the droplets from the orifice. The nozzle or orifice is often easily replaceable because many inks can cause nozzle erosion which can alter the droplet size.

Inkjet printing, regardless of the specific printing regime or format, is largely dependent on droplet size and spreading dynamics. Flexible organic circuits and arrays, especially those including OFETs or OLEDs, are well suited to inkjet printing since the semiconducting polymers and insulating polymers such as the cyclotenes are especially amenable to droplet formation and deposition onto substrates, often through heating of the inkjet head. Inkjet droplets are measured in picoliters (10^{-12} l). Typical photo inkjet printers eject drops having volumes between 1 and 4 picoliters, corresponding to drop diameters ranging from less than 10 μm to less than 100 μm. A 1 picoliter (pL) drop could contain roughly 1 million nanoparticles 10 nm in diameter. Because of the small drop size, the drops are spherical and controlled by surface tension. These small drop sizes can translate into smaller or larger ink pattern sizes depending on the absorption or wetting of the substrate. The deposition of single-component droplets such as specific polymer materials differs from more conventional inks in that the droplet serves as a matrix or binder for the ink particulates, which for electronic printing can be inorganic particles such as nanoparticles or even nanorods, especially for the conducting metal contacts or interconnect wiring.

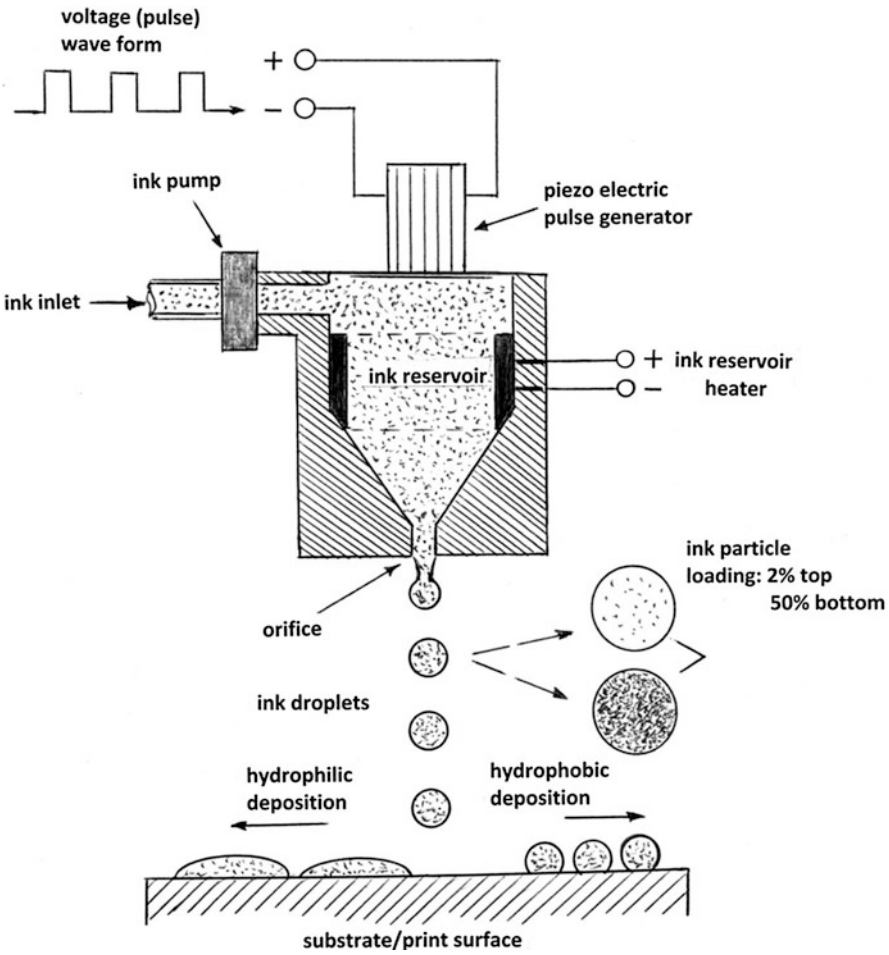


Fig. 6 Drop-on-demand inkjet schematic. Printer head has piezo generator for drop release and reservoir heater to heat ink. The substrate can also be heated as necessary to anneal or sinter ink particles, especially metal nanoparticle conductor interconnects and device contacts. Different droplet loadings are also illustrated

Figure 6 illustrates schematically a piezoelectric (piezo) drop-on-demand inkjet print head printing on a wetting (hydrophilic) surface to the left and a relatively non-wetting (hydrophobic) surface to the right. Printing 3D structures such as organ components illustrated conceptually in ▶ Fig. 4 of chapter "Tissue Engineering Scaffolds and Scaffold Materials" as well as other 3D structures such as 3D electronic structures requires repeated passes of the printer head, and the layer deposition is then not influenced by the substrate. However, for thin-layer interconnects, the droplets would overlap, and the binder would be removed by heating.

For inkjet deposition of metal conductors, Cu, Ag, Au, and Pt have been used along with Al, although Al represents the poorest conductor. Particle size is critical since the most desirable circumstances are those in which the binder can be evaporated while the metal particles sinter (or melt) together. But in bulk form, even Al melts well above the temperature which could be tolerated by 3D flexible electronic fabrication, especially polymeric devices. However, it has long been known that melting point is suppressed for nanoparticle materials of all types, including ceramic and metal nanoparticles. This melting point suppression declines with decreasing particle size as represented generally by Buffat and Borel (1976)

$$\left(T_b - T_p\right) = \left[\frac{2T_b}{\Delta H\, \rho_s\, r_s}\right]\left[\gamma_{SV} - \gamma_{LV}\left(\frac{\rho_s}{\rho_L}\right)\right]^{2/3} \tag{1}$$

where T_b and T_p represent the bulk melting point and particle (especially nanoparticle) melting point, Δ H is the molar latent head of fusion, ρ_s and ρ_L are the solid and liquid phase density, r_s is the particle radius, and γ_{SV} and γ_{LV} represent the solid-vapor surface free energy and liquid-vapor surface free energy (or surface tension), respectively.

Figure 7 illustrates the melting point decrease for nanoparticle size decrease of gold and silver. Silver nanoparticles in the range of 5 nm have been shown to have a melting temperature ~140 °C, and these nanoparticles, like Au of similar size (Fig. 7), can be produced by chemical reduction of the corresponding metal salts:

$$2AgNO_3 + 2NaBH_4 \rightarrow 2\mathbf{Ag} + H_2 \uparrow + B_2H_6 + 2\,NaNo_3 \tag{2}$$

$$AuCl_4^- + 2\overline{e} \rightarrow AuCl_2^- + 2\,Cl^- \qquad AuCl_2^- + \overline{e} \rightarrow \mathbf{Au} + 2\,Cl^- \tag{3}$$

Copper nanoparticle ink can be more easily prepared by copper oxide reduction and will respond similar to Au and Ag in Fig. 7. Copper is considerably cheaper than Ag, and this is an advantage in high-volume production processes.

Considering Fig. 7 in the context of Fig. 5 in retrospect, it is necessary to reexamine the inkjet and especially ink technology more broadly. The ink or ultimately the droplet created by the pulsed printer head in Fig. 5 is a complex regime. Most importantly, the ink particles (or nanoparticles) must be uniformly dispersed in the matrix or binder. This may involve surface treatment such as the attachment of organic ligands, etc. Depending upon the deposition function, the particles (ink particles) can be nanometals such as Ag or Au in Fig. 7, oxide nanoparticles (which may also follow a melting point suppression illustrated in Fig. 7), semiconductor particles such as CdS, GaN, etc.; or carbon nanoparticles and carbon nanotubes. Multiple printer heads containing functional polymeric materials which can be heated to establish requisite flowability, viscosity, surface tension, deposition wetting, etc., can be combined in printing arrays to fabricate hybrid electronic or microelectronic circuits, etc., or devices implicit in Fig. 4.

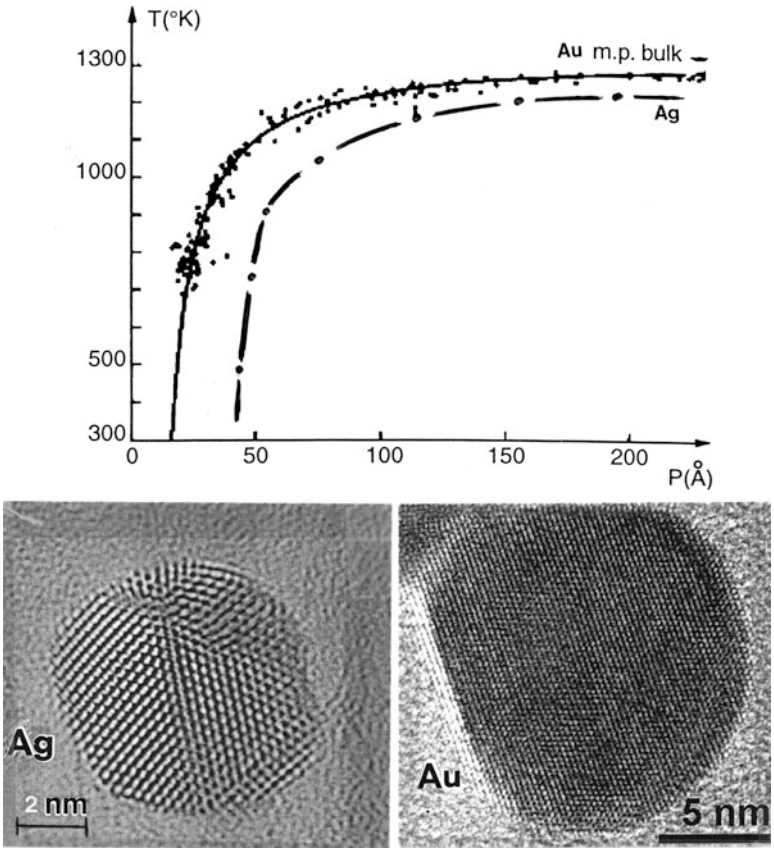


Fig. 7 Melting point suppression for Au and Ag nanoparticles. Solid line for Au represents a least-square fit to Eq. 1, while the solid-dotted line represents a least-squares fit to Eq. 1 for Ag. The experimental data (dots) for Au is the work of Buffat and Borel (1976). The TEM atomic resolution image is a silver nanoparticle adapted from Scholl et al. (2012). TEM atomic resolution Au image is similar to Ag. In the graph, the particle diameter (in Angstrom units) is indicated ρ(Å)

In many ink (or ink droplet) compositions, the matrix or binder must be formulated for a requisite rheology or viscosity and surface tension in relation to ink particle or nanoparticle loading (Fig. 6) and the impact of drop size on spreading (or wetting) dynamics on different substrates and the role of nozzle size. The removal of the binder and optimization of conductivity through nanoparticle melting, etc., is also important in controlling functionality of printed circuits and device components.

Both viscosity and surface tension for most liquids and liquid polymers will decrease with temperature. Surface tension of single-component liquids such as water or alcohol will follow a relationship illustrated by Eq. 205 in chapter "► A Brief Introduction to Quantum Mechanics." However, water has a high surface tension (~73 mJ/m^2) and low viscosity, while alcohol (ethanol) has a low

surface tension (~43 mJ/m^2) relative to water and a correspondingly low viscosity. Viscosity as described by Newton's law of viscous flow

$$F = \eta A\left(\frac{dV_x}{dz}\right) \tag{4}$$

implies that if the viscosity, η, is constant for a liquid (fluid), it is referred to as Newtonian, while if η is not constant, it is non-Newtonian. Polymer solutions or melts, blood, saliva, etc., are non-Newtonian fluids.

Since viscosity and surface tension for fluids generally decreases with increasing temperature, it might be expected that they are related. However, viscosity, η, measured in Poise or centipoise (cP), measures the resistance to flow or internal friction of a fluid, while surface tension, γ_{LV}, relates to the ability of the liquid surface (liquid/vapor interface, LV) to resist a force (or deformation). Pelofsky (1966) related these thermophysical properties by the relation:

$$\ell n\gamma_{LV} = \ell n\, A + B/\eta \tag{5}$$

However, the experimental constants A and B in Eq. 5 must be considered for each specific liquid, and no simple and generally applicable relationship has been observed. The viscosity can also be described by an equation of the form

$$\ell n\ \eta = \ell n\, C + E_a/RT, \tag{6}$$

where C is another specific fluid constant, E_a is the energy to dissociate the liquid molecules, R is the gas constant, and T is the temperature. A plot of $\ell n\ \eta$ versus 1/T will give E_a/R.

The natural logarithm of viscosity ($\ell n\ \eta$) will also generally vary in direct proportion to particle loadings in weight percent as illustrated in Fig. 6. In the case of metal nanoparticles dispersed in purified (deionized) water, a viscosifier such as hydroxyethyl cellulose (HEC) can be added along with some appropriate surfactant. Conductor printing can involve loadings of 1–70 % (weight) nanosilver, for example, with surface tensions ranging from 20 to 40 mJ/m^2 for solvent-based binders and 30–60 mJ/m^2 for water-based dispersions. Corresponding viscosities range from 1 to 200 cP at room temperature, but in order to produce continuous, melted silver conductors or conducting interconnect strips, the binder can be heated to ~60 °C, while the substrate can be heated to ~140 °C for <5 nm Ag particle loadings. Normally this temperature would also simultaneously evaporate the binder solvents but not compromise other device integrity or melt the substrate material. Common binder solvents have been composed of di(propylene glycol) methyl ether (DPM) or butoxyethyl acetate (BEA). Sharp detail is often achieved by droplets as small as 6 pL with transition up to 36 pL for more uniform deposition. This is achieved by changing the orifice or having printer heads with varying orifices.

A liquid drop on a substrate can be considered a sessile drop as illustrated in Fig. 8 where the contact angle (as in ▶ Fig. 9 of chapter "Biomimetics and

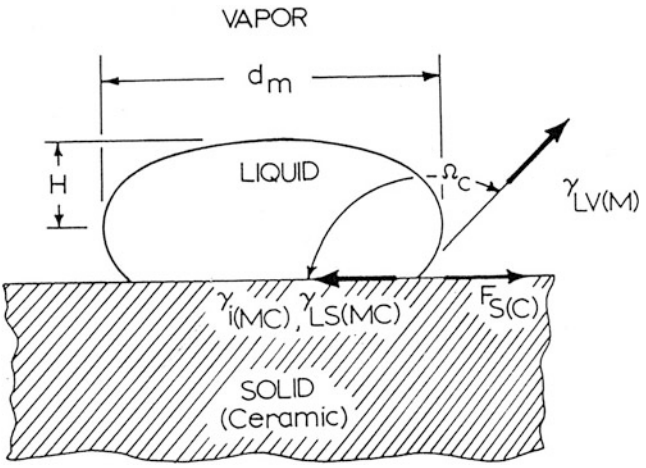


Fig. 8 Liquid (sessile) drop on solid substrate. The contact angle is denoted $\Omega_c \cdot \gamma$; (MC) = γ_{LS} (MC), the liquid/solid interfacial free energy

Biologically Inspired Materials") is denoted Ω_c, and the equilibrium of surface tension and specific surface free energies of the substrate (F_{scc}) and the interfacial free energy for the liquid drop/substrate interface, γ_{LS}, is given by the Young equation

$$\gamma_{LS} = \gamma_{LV} \cos(180 - \Omega_c) + F_{scc} \tag{7}$$

where γ_{LV} the surface tension for the liquid drop is given approximately by

$$\gamma_{LV} \cong \frac{g\rho_L dm^2}{4}\left(\frac{0.052}{f} - 0.123 + 0.48\, f\right), \tag{8}$$

where g is the gravitational force (or constant) (g = 9.8 m/s^2), ρ_L is the liquid drop density at temperature on the substrate, d_m is the maximum drop diameter as shown in Fig. 8, and

$$f = (2\,H/d_m) - 0.4l4, \tag{9}$$

where H is the drop height measured from the maximum diameter line (d_m). Equation 8 is particularly applicable to small drops (Murr 1975), although the smaller the drop, the more the chemistry (Magdassi 2009) or composition, including the loading of the binder (matrix) with nanoparticles in the case of nanoparticulate inks.

The adhesion of a liquid drop on a substrate may also be important and the so-called reversible work of adhesion (W_{ad}) is expressed as

$$W_{ad} = \gamma_{LV}(1 + \cos\Omega_c) \tag{10}$$

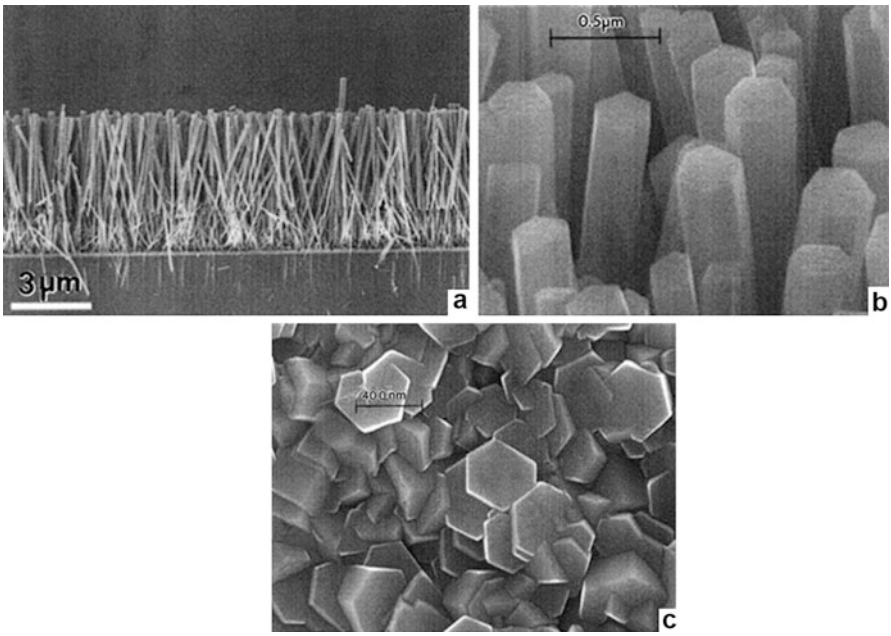


Fig. 9 ZnO nanorod and nanocrystal growth on ITO substrates. (**a**) SEM of oriented ZnO nanorods. (**b**) Magnified view of ZnO nanorods in (**b**). (**c**) Hexagonal ZnO nanocrystals. (**a**) and (**b**) are adapted from Wang et al. (2009). (**c**) is adapted from Chae et al. (2010)

Correspondingly, wetting or hydrophilicity improves practical adhesion, at least at the time of drop deposition on a substrate.

Of course it is important to prevent orifice clogging in the printer head (Fig. 6), and this may be important in printing nanorods or nanotubes, especially where the intent is to orient them as implicit in Fig. 5, for example. There may also be some advantage in mixtures or distributions of nanoparticles or nanoparticle shapes. Figure 9 illustrates the growth of aligned, hexagonal, ZnO nanorods along with ZnO hexagonal nanocrystals grown on a ZnO – seeded ITO (indium tin oxide) substrate at 60 °C (Wang et al. 2009; Chae et al. 2010). In contrast, Figs. 10 and 11 illustrate shapes and shape evolution for Au nanoparticles which can also be compared with the more equifaceted or spherical-related Au nanoparticle shapes as shown in Fig. 7. The pentagram-shaped Au nanorods in Fig. 11b represent a novel quasicrystal nanorod form grown in a bath of silver nitrate, ascorbic acid, and gold chloride. This silver-assisted, seed-mediated growth alters the faceting and facet structure as this relates to shape evolution (Personick et al. 2011; Langille et al. 2010). The more regular Au nanorods in Fig. 11a grow in environments where the ascorbic acid used to grow the quasicrystal rods in Fig. 11b is replaced by hydroquinone (Vigdermon and Zubarev 2012).

It might be noted that gold nanorods such as those illustrated in Fig. 11a have shown promise in several medical applications. Prominent among them is the

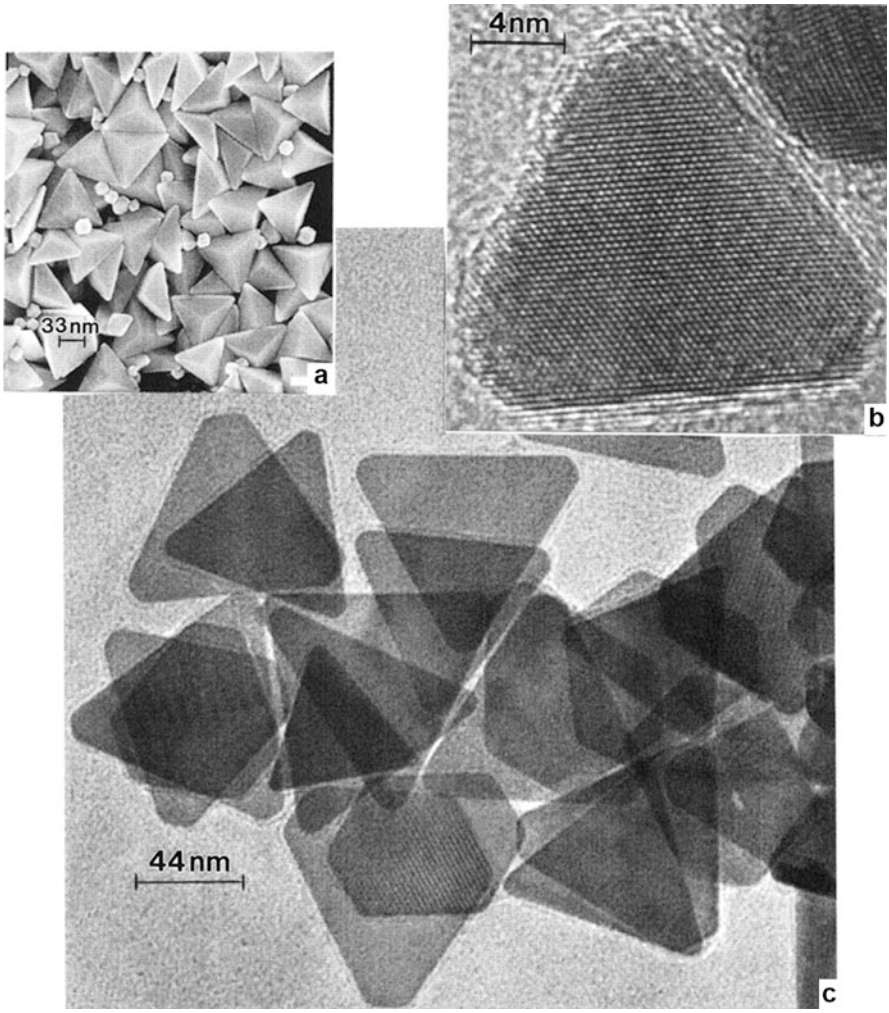


Fig. 10 Au bipyramids (**a**) and {111}-faceted octahedra (**b**) and (**c**). (**a**) is adapted from Personick et al. (2011), (**b**) shows an atomic resolution TEM image courtesy of Beatriz Rodan Cuenya, University of Central Florida, and (**c**) shows a TEM image adapted from Jones et al. (2010)

vectoring of these nanorods to cancer cells where they selectively aggregate. Such aggregation can allow for contrast for in vivo and in vitro imaging as well as a regime for focused heating using microwave therapies to selectively kill cancer cells. Nanorods differ from the more traditional spherical-like gold nanoparticles illustrated in Fig. 7 by their absorption of infrared (IR) radiation. This can allow an IR laser to activate them without damaging surrounding cells. In some cases, the gold surface must be functionalized as indicated previously for other nanoparticle functionalization.

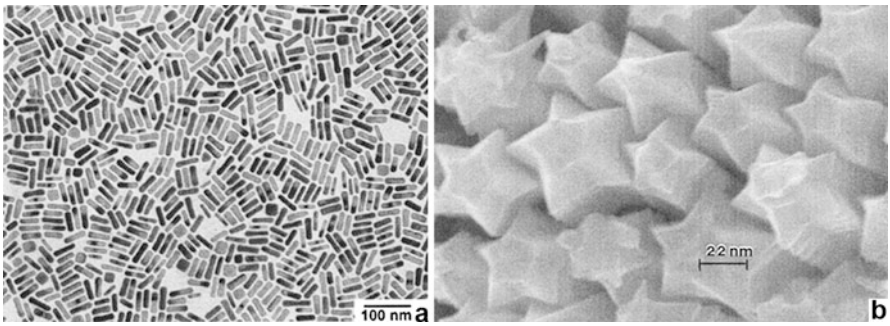


Fig. 11 Gold nanorods (**a**) and quasicrystal (pentagram)-shaped gold nanorods (**b**). (**a**) is courtesy of Kimberly Hanad-Schifferli, MIT. (**b**) is courtesy of Eugene Zubarev, Rice University

Huang et al. (2013) have also described an artificial protein composite containing Au nanorods (∼15 nm diameter and 50 nm in length) which is used as a novel glue to help "weld" tissue ruptures by absorbing near-IR laser light which passes harmlessly through human tissue, but causing the Au nanorods to heat up as a consequence of the absorbed energy, coagulating the protein matrix. Such laser applications are used in tissue repair in cartilage blood vessels, urinary tracts, and tissue repair in inflammatory bowel disease, which afflicts ∼1.5 million people annually in the USA. The protein-based, Au nanorod glue exhibits potential in patching holes in weak colon tissue.

Printer configurations can take many forms depending upon the application. 3D printing for multilevel device structures and printing on irregular surfaces requires variances in the height of the print head, travel velocity, print resolution, accuracy and repeatability of patterns or structures, and the like. Some applications employ continuous dispensing of the ink, and complex structure printing can employ multiple dispensing heads with a wide range of viscosities. In addition, hybrid fabrication processes can integrate stereolithography with direct print technologies as we will discuss later.

References

Buffat P, Borel J-P (1976) Size effect on melting temperature of gold particles. Phys Rev A13:2287–2291

Chae K-W, Zhang Q, Kim JS, Jeong Y-H, Jeong Y-H, Cao G (2010) Low temperature solution growth of ZnO nanotube arrays. Beilstein J Nanotechnol 1:128–134

Chun J, Hwang Y, Choi Y-S, Jeong T, Baek JH, Ko HC, Park S-J (2012) Transfer of GaN LEDs from sapphire to flexible substrates by laser lift-off and contact printing. IEEE Photon Technol Lett 24(23):2115–2118

Huang H-C, Walker CR, Nanda A, Rege K (2013) Laser welding of ruptured intestinal tissue using plasmonic polypeptide nanocomposite solders. ACS Nano 7(4):2988–2998

Jones MR, MacFarlane RJ, Lee B, Zhang J, Young KL, Jenes AJ, Mirkin CA (2010) DNA-nanoparticle superlattices formed from anisotropic building blocks. Nat Mater Lett 9:913–917

Klauk H (2010) Organic thin-film transistors. Chem Soc Rev 39:2643–2652
Langille MR, Personick MC, Zhang J, Mirkin CA (2010) Defining rules for the shape evolution of gold nanoparticles. J Am Chem Soc 134:14542–14544
Lunt RR (2012) Theoretical limits for visibly transparent photovoltaics. Appl Phys Lett 101:043902 (1–3)
MacDiarmid AG (2001) Synthetic metals: a novel role for organic polymers. Angew Chem 40:2581–2590
Magdassi S (2009) The chemistry of inkjet inks. World Scientific, Singapore
Murr LE (1975) Interfacial phenomena in metals and alloys. Addison-Wesley, Reading
Nyström G, Razaq A, Stromme M, Nyholan L, Mihranyan A (2009) Ultrafast all-polymer paper-based batteries. Nano Lett 9(10):3635–3639
Pelofsky AH (1966) Study of surface tension and viscosity for mixed alkanes. J Chem Eng Data 11:394–397
Personick ML, Langille MR, Zhang J, Harris N, Schatz GC, Mirkin CA (2011) Synthesis and isolation of {110}-faceted gold bipyramids and rhombic do deca hedra. J Am Chem Soc 133:6170–6173
Repo P, Haarahiltunen A, Sainiemi L, Yli-Kooki M, Talvitie H, Shubert MC, Savin H (2013) Effective passivation of black silicon surfaces by atomic layer deposition. IEEE J Photovolt 3(1):90–95
Scholl JA, Koh AL, Dionne JA (2012) Quantum plasmon resonances of individual metallic nanoparticles. Nature 483:421–427
Vigdermon L, Zubarev ER (2012) High-yield synthesis of gold nanorods with longitudinal SPR peak greater than 1200 nm using hydroquinone as a reducing agent. Chem Mater 25:1450–1457
Wang Y, Yeow JTW, Chen L-Y (2009) Part 10, Synthesis of aligned ZnO nanorods with different parameters and their effects on humidity sensing. In: Murthopadhyay SC, Gupta GS, Huang RY-M (eds) Recent advances in sensing technology, vol 49, Lecture notes in electrical engineering. Springer, New York

Bioprinting and Biofabrication of Organs

Contents

Abstract

Using living cells in place of nanoparticles of other materials in an ink matrix, it is conceptually possible to print 3D organs. Using ink jet arrays controlled by CAD software, vascular cells can introduce appropriate vascularization or cylindrical vessels to allow blood flow. The successful integration of these functions into tissue engineering can allow for human organ printing. This chapter describes contemporary innovations in tissue engineering and 3D organ printing.

Introduction

As noted in connection with ▶ Fig. 4 of chapter "Tissue Engineering Scaffolds and Scaffold Materials," one of the ultimate accomplishments in the widest sense of materials science and engineering would be the systematic selection and organization of living cells in the fabrication of tissue and functional organs. Commonly referred to as tissue engineering which has been directed toward the use of biodegradable scaffolds as partial support for the assembly of isolated cells into contiguous tissue (Ma and Elisseeff 2005; Langer 2007). However, the ultimate goal of tissue engineering involves the design and fabrication of functional human organs. These are technically "artificial" organs which would be suitable for tissue or organ regeneration or replacement. As noted in chapter "▶ Implant Materials and Structures," any replacements in the body must be biocompatible. These include artificial tissue, organs, prostheses, or other devices or appliances.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_36

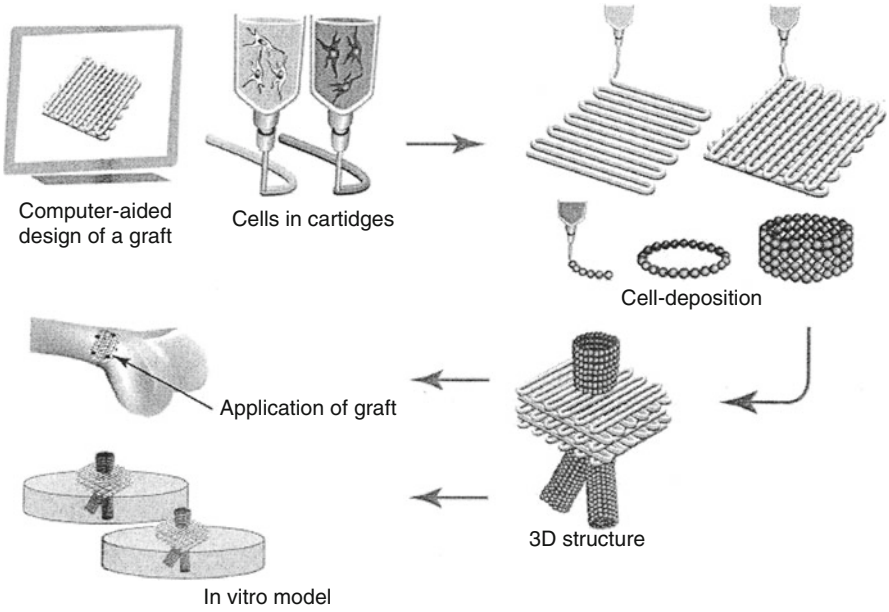


Fig. 1 Bone graft organ/tissue printing using multiple printing heads containing specific cell types suspended in a hydrogel (after Bello et al. 2001). The printer heads can deposit continuous filaments or drops as appropriate

By considering cell speciation and function in the self-assembly fabrication of tissue, it is possible to mimic the process by printing tissue and ultimately organ structures by embedding these structures, composed of speciated cells, cell matrices, and cell geometries and spatial configurations in computer-aided design (CAD) models which would direct cell deposition from dispensing heads (orifices) or inkjet printer head configurations as shown schematically in ▶ Fig. 6 in chapter "3D Printing: Printed Electronics". Such bioprinters, with multiple dispensing heads corresponding to specific cells within a suitable, biological binder or droplet could print tissue or organ structures layer by layer in 3D printing. Figure 1 illustrates this concept schematically for the heterogeneous printing of bone grafts which can form bone tissue once implanted (Bello et al. 2001).

Organ Printing

The challenge in organ printing is vascularization of the tissue since without blood flow to provide oxygen, cells will die. Consequently, either vascular-specific cells must self-assemble or self-organize among other tissue-specific cells, such as osteoblasts in bone regeneration components (Fig. 1) or fibroblasts in other tissue construction. The suffix "blast" implies stem cells on a cell in an activated state of metabolism or function.

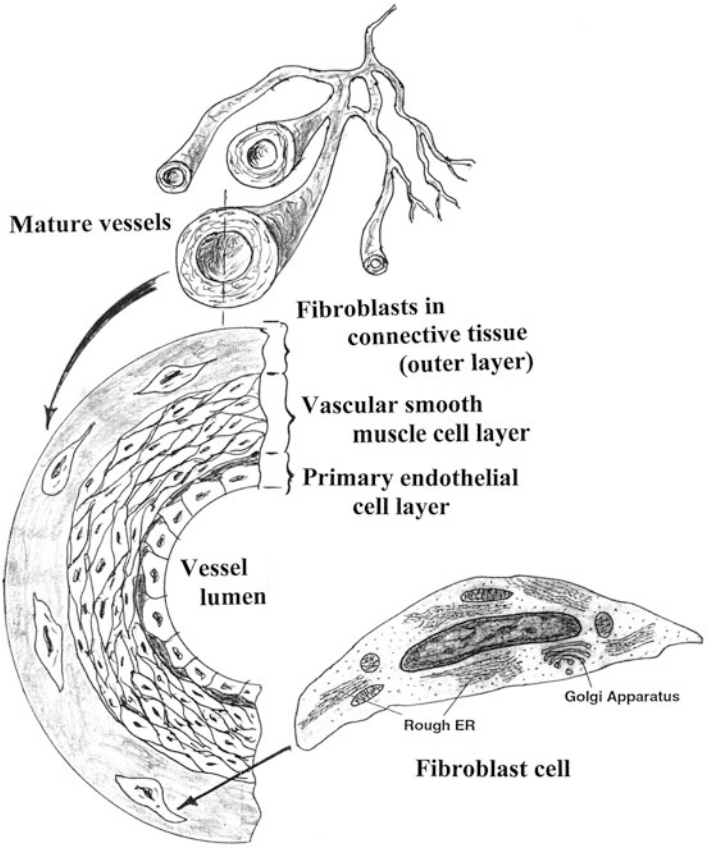


Fig. 2 Blood vessel development and structure

The smallest blood vessels are single-layer capillaries composed of endothelial cells, which originate from angioblasts, and the most primitive have a single endothelial cell wrapped around to join with itself, forming a primitive cell ring. These capillaries can grow or self-assemble, and as they grow they add layers to thicken the cell walls forming vascular tissue. With increasing diameters and wall thickness, these capillaries bifurcate, forming a fractal-like hierarchy (vascular tree) illustrated for lungs and trees in ▶ Fig. 12 of chapter "Biomimetics and Biologically Inspired Materials." Figure 2 illustrates the evolution of mature blood vessels and their structure. The outer layer for these mature vessels is composed of fibroblast cells in a fibrous extracellular, protein matrix, separated from the inner endothelial cell layers by a vascular smooth muscle cell layer as shown in ▶ Fig. 2 of chapter "Bioprinting and Biofabrication of Organ." The fibroblast cells in the enlarged cartoon in Fig. 2 contain rough endoplasmic reticulum (ER) inactive cells which synthesize proteins. The Golgi apparatus

modifies, sorts, and packages these macromolecular proteins for use within the cell or for cell secretion to the supporting matrix or connective tissue. The most prominent fiber type in connective tissue is collagen as illustrated in the hierarchical structure in ▶ Fig. 2 of chapter "Examples of Natural Composites and Composite Structures." Connective tissue also contains fibroblasts whose stem cell line also produces osteoblasts of bone, adipocytes of fat, and chondrocytes in cartilage tissue. Fibroblasts can regress back to earlier stages in their development and differentiate into osteoblasts to aid bone production or remediation or chondrocytes and adipose cells under some circumstances. Fibrocytes are sometimes referred to as cells which have become quiescent, maintenance cells rather than production cells. Fibroblasts become active or productive in normal growth and wound repair. The flat fibroblast cells shown in Fig. 2 actually lie on top of or along the fibers they produce in connective tissue. During the healing process, fibrin, a fibrous, non-globular protein, is formed from fibrinogen by the protease thrombin and polymerized to form a mesh-like structure. It is also involved in blood clotting. The fibrin matrix allows migration and attachment of neutrophils and macrophage cells which are among the most notable protector cells in organ tissue.

The largest organ in the human body is skin which acts as the containment regime, protecting the body from pathogens, water loss, and thermal management. Skin has the ability to heal naturally. It is composed of two layers, the epidermis (the outer layer) and the dermis (the inner layer). The principal cells in the epidermis are keratinocytes which manufacture keratin (▶ Figs. 2 and ▶ 16 of chapter "Structures and Properties of Keratin-Based and Related Biological Materials") to maintain the mechanical integrity of the epidermal tissue regime. The dermis also has an underlying subcutaneous tissue layer or hypodermis which is largely adipose (or flat) tissue. The dermis supports the epidermis by providing nutrients and oxygen and therefore contains vascularized tissue. The principal proteins composing the connecting tissue are collagen, elastin (▶ Fig. 19 of chapter "Structures and Properties of Keratin-Based and Related Biological Materials"), and glycosaminoglycans. Fibroblast cells also populate the dermis and especially around blood vessels (Fig. 2).

Because skin is a shallow organ with a very large surface area, it has been one of the first bioprinting efforts. The fabrication or biofabrication of skin (or artificial skin) also has significant practical implications and applications since nearly half a million people are treated for burn injuries annually in the United States, while diabetes-related skin treatments exceed this number by more than an order of magnitude. Tissue engineering of skin and inkjet printing or bioprinting of skin have been pursued over the past decade or more with improving success (Bello et al. 2001; Boland et al. 2006; Cui and Boland 2009; Bottcher-Haberzeth et al. 2010). Especially notable has been the formation of capillary structures in 3D with human dermal microvasculature endothelial cells in a collagen or fibrin matrix (Cui and Boland 2009; Bottcher-Haberzeth et al. 2010).

Fig. 3 3D-printed, thin artificial skin (Google image)

Most commercially available artificial skin serves as a wound covering which facilitates wound healing and limits contraction during healing. The first artificial skin made from human cells was a product called Dermagraft-TC, approved by the US FDA in 1997. Figure 3 illustrates a 3D-printed, thin human cell artificial skin covering. These artificial skin products are too thin to accommodate vascularization. Recent efforts to print cells directly onto a wound site to support natural cell regeneration and wound healing have been pursued. The use of patient fat cells (adipocytes) to print customized tissue configurations such as breast reconstruction has also been studied (Melchelo et al. 2011).

Yanez et al. (2012) have recently investigated a 3-layer, 3D-printed skin grafts by printing the dermal layer onto a biopaper (biomimetic hydrogel) using a collagen bioink (hydrogel matrix) with fibroblast cells, an intermediate layer using a fibrin gel bioink containing endothelial cells, and an outer or epidermal layer using a collagen-based hydrogel bioink containing keratinocyte cells. The strategy in this approach was to encourage vascularization in the intermediate layer printed with endothelial cells (Fig. 2). This printing strategy can be modified by pointing networks of strings of endothelial cells within the dermal layer from a separate dispersing nozzle to allow the cells to form perfusion capillaries by self-assembly and self-organization.

In the more general case of organ printing as a 3D-additive manufacturing process, Mironov et al. (2003) identified three sequential steps. The first involves the development of a CAD design or blueprint of a specific organ. This can be derived from digital image reconstruction such as microcomputer tomography (CT) scans, MRI, or detailed reconstructions of serial sections (Sun and Lai 2002). These CAD models can be embedded in stereolithographic (STL) files to

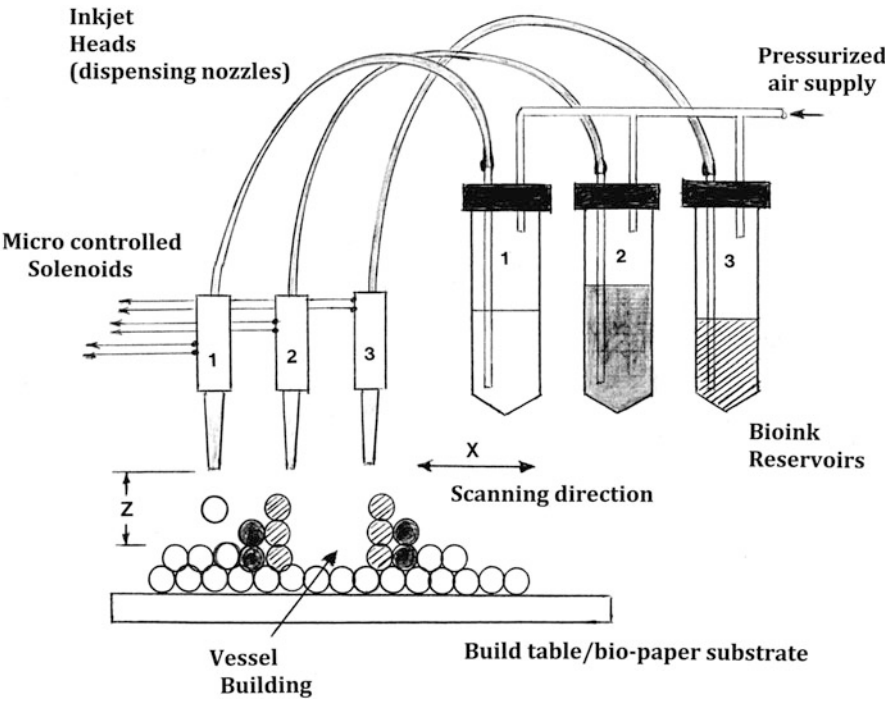


Fig. 4 Schematic section showing conceptual 3D blood vessel building using multiple dispensing heads for specific tissue spheroid printing. (*1*) Fibroblast cells in fibrin cell hydrogel; (*2*) fibroblast + smooth muscle cell hydrogen. (*3*) Endothelial cells in hydrogel. Hydrogel is collagen based. Z represents the dispensing height which is constant with layer building

direct the printer heads in specific bioink deposition in each layer. This is illustrated schematically in Fig. 4 for the printing of a simple, mature blood vessel. The implications of Fig. 4 are that tissue spheroids as bioinks are the building blocks in organ printing as originally proposed by Mironov et al. (2009) for continuous or digital (drop-on-demand) printing from multiple dispensing heads each printing functional tissue or organ components. This involves the preparation of cell compositions and extracellular material along with growth factors and other critical growth compositions (molecular structures) to form controlled spheroids having requisite surface tension and viscosity as discussed in connection with 3D inkjet printing illustrated in ▶ Fig. 6 of chapter "3D Printing: Printed Electronics" in order to assure tissue fusion. This may also involve maintaining an optimum temperature in the printer reservoir as illustrated in ▶ Fig. 6 in chapter "3D Printing: Printed Electronics".

Figure 5 shows a roadmap for printing microvascular segments of a vascular tree as developed by Mironov et al. (2009). The vascular networks are assumed to self-assemble from endothelial tissue spheroids implicit in the schematic of

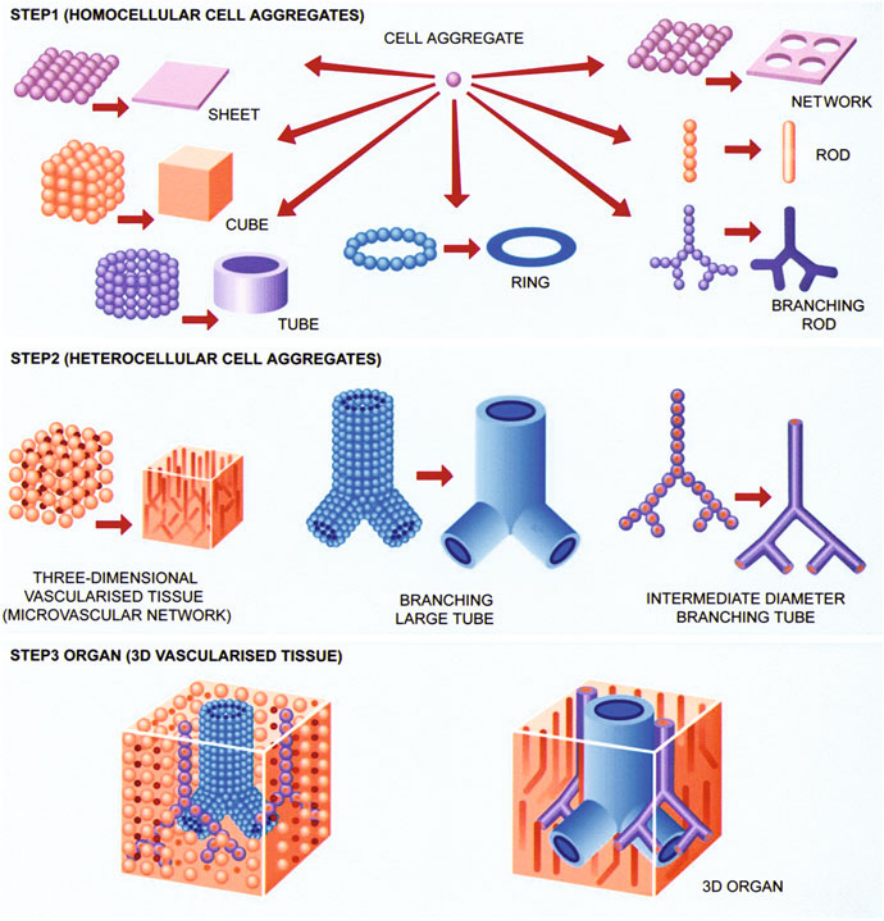


Fig. 5 3-step model roadmap for 3D organ printing (From Mironov et al. (2009)

Fig. 4. Most organs are prominently represented by a vascular tree which, as illustrated for the kidney in Fig. 6, can be embedded in a CAD model along with associated STL files for connective or supporting tissue fabrication in layer-by-layer printing.

The kidney, as illustrated in Fig. 6, functions as a blood filter, removing wastes and excess water which are diverted to the urinary bladder through the ureter. The kidney receives blood from the renal artery which branches directly from the abdominal aorta. Correspondingly, the two kidneys receive roughly 20 % of the cardiac output or continuous blood flow. Since kidney dialysis has been a viable intervention for kidney function, kidney fabrication for implantation might be

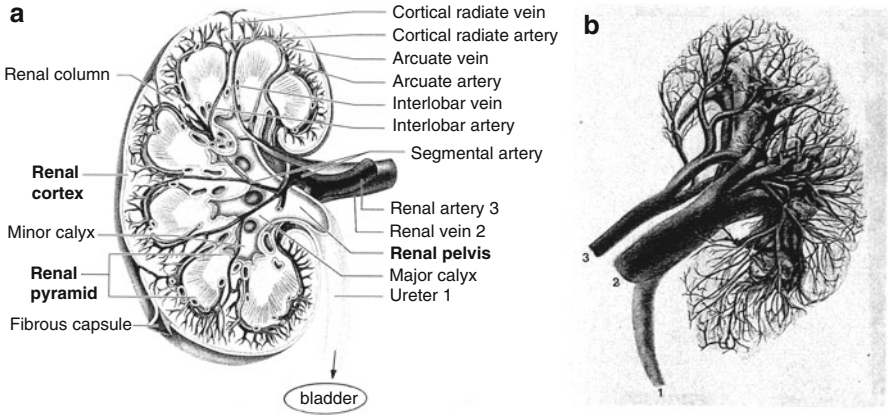


Fig. 6 Kidney structure and function. (**a**) Kidney structure showing the complete organ cutaway. (**b**) Vascular and renal free structure of the kidney. The major vessel components indicated by 1, 2, and 3 correspond to the ureter (*1*), renal vein (*2*), and renal artery (*3*) in (**a**)

simplified by incorporating the engineering design principles into the CAD models for 3D bioprinting. In addition, the printing of kidneys and other organs, even skin, can utilize the patient's own cells or stem cells speciated with the patients DNA. This would eliminate implant rejection.

In 1990, there were 4,000 kidney donations from deceased donors in the United States and 17,000 on the kidney waiting list. These numbers jumped to 7,000 donors and nearly 60,000 on the waiting list in 2003. Projections for 2014 are doubled. Consequently organ manufacturing can make a huge contribution.

Figure 7 illustrates a commercial, 3D-bioprinting system. At some point, when it becomes feasible to manufacture patient-specific organs, organ-specific manufacturing stations could conceivably occupy mall space, where medical practitioners could order organ fabrication for a patient by transmitting speciated cells and micro-CT data for the organ to be replaced. This could of course apply to skin grafts or replacements as well where the physician could order a specific surface area to be manufactured.

To date (2013), experimental knee cartilage, heart valves, bone implants, and skin have been printed. Organovo, Inc. has made blood vessels and lung tissue. Pediatric ears have been fabricated from cartilage tissue using a collagen hydrogel scaffold and respective CAD models for molds precisely mimicking the normal anatomy of patient-specific ears (Reiffel et al. 2013).

It should be apparent that in the biological context of organisms and organs, cells play a central role as a consequence of their selectivity, or more specifically their structure and function. However, cells continually remodel their internal structure, thereby altering their mechanical properties (Bausch and Kroy 2006). Indeed, in a recent paper, Karen Kasza and colleagues discuss the cell as a material, if not a

Fig. 7 Commercial, 3D bioprinter manufactured by nScript, Inc., and housed in the bioengineering laboratories at the University of Missouri

living material (Kasza et al. 2007). Consequently, and as discussed by Vladimir Mironov and colleagues (2009), tissue spheroids can also be considered a living material with measurable and controllable properties, e.g., surface tension and viscosity.

References

Bausch AR, Kroy K (2006) A bottom-up approach to cell mechanics. Nat Phys 2:231–238

Bello Y, Falabella A, Eaglstein W (2001) Tissue-engineered skin. Current status in wound healing. Am J Clin Dermatol 2(5):305–313

Boland T, Xu T, Damon B, Cui X (2006) Application of inkjet printing to tissue engineering. Biotechnology 1:910–917

Bottcher-Haberzeth S, Biedermann T, Reichmann E (2010) Tissue engineering of skin. Burns 36:450–460

Cui X, Boland T (2009) Human microvasculature fabrication using thermal inkjet printing technology. Biomaterials 30:6221–6227

Kasza KE, Rowat AC, Liu J, Angelini TE, Brongwynne CP, Koenderink GH, Weitz DA (2007) The cell as a material. Curr Opin Cell Biol 19:101–107

Langer R (2007) Tissue engineering: perspectives, challenges, and future directions. Tissue Eng 13:1–2

Ma PX, Elisseeff J (eds) (2005) Scaffolding in tissue engineering. CRC Press, Boca Raton

Melchelo F, Domingos PW, Neves MA, Klein TJ, Maida J, Bartolo PJ, Hatmacher D (2011) Additive manufacturing of tissue and organs. Prog Polym Sci 37(31):1079–1104

Mironov V, Boland T, Trask T, Forgacs G, Markwald RR (2003) Organ printing: computer-aided jet-based 3D tissue engineering. Trends Biotechnol 21:157–161

Mironov V, Visconti RP, Kasyanov V, Forgacs G, Drake CJ, Markwald RP (2009) Organ printing: tissue spheroids as building blocks. Biomaterials 30:2164–2174

Reiffel AJ, Kafka C, Hernandez KA, Popa S, Perez JL, Zhau S, Pramanik S, Brown BN, Ryu WS, Bonassar LJ, Spector JA (2013) High-fidelity tissue engineering of patient-specific auricles for reconstruction of pediatric microtia and other auricular deformities. PLoS One 8(2):e56506

Sun W, Lai P (2002) Recent development on computer-aided tissue engineering – a review. Comput Methods Programs Biomed 67:85–103

Yanez M, Rincon J, Cortez P, Gänthes N, Boland T (2012) Printable cellular scaffold using self-crosslinking agents. J Imaging Sci Technol 56(4):040506-1–040506-5

Rapid Prototyping Technologies: Solid Freeform Fabrication

Contents

Abstract

Rapid prototyping or solid freeform fabrication fundamentals and processes are described in this chapter. These include concepts of layer printing to produce 3D objects by fused deposition modeling and selective laser melting. Stereolithography as it is applied to fabrication by additive manufacturing forms the basis for these technologies for materials processing.

Introduction

Rapid prototyping (RP) involving solid freeform fabrication (SFF) represents a range of additive (layer) manufacturing concepts where a computer-generated model or computer-aided design (CAD) file format is converted (or decomposed) into individual, oriented, horizontal layers ~100–250 μm thick. These layers are often referred to as 2-1/2 dimensional information regimes. During fabrication, each layer or slice is physically deposited or selectively created by photosensitive liquid resin curing in a vat: photopolymer-based processing. Although there are more than two dozen variances of these 3D-printing technologies, many are based on three major methodologies: stereolithography (SL, STL, SLA, SL-RP, etc.), fused deposition modeling (FDM), and (selective) laser sintering (SLS or LS) (Jacobs 1992; Gibson et al. 2010; Bartolo 2011).

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_37

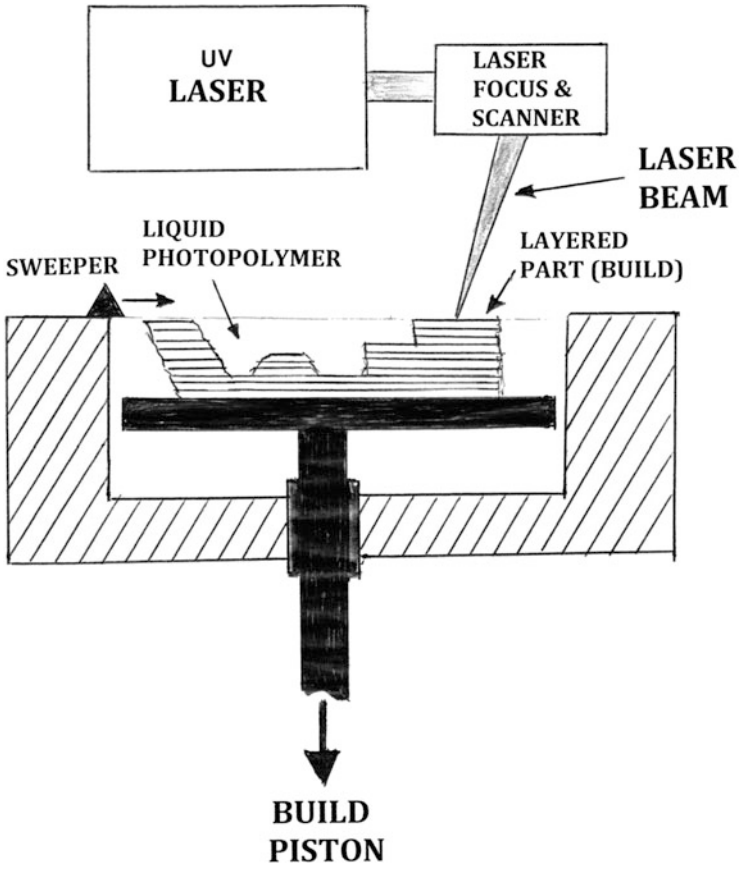


Fig. 1 Stereolithography system schematic, with each photo (laser) – cured layer or layer portion, the sweeper removes excess photopolymer or unreacted polymer and the build piston moves the build table down a predetermined layer distance. The laser beam is focused through glass lenses and scanned using a mirror driven by the CAD (STL) file, which includes a blanking system

Stereolithography: 3D-CAD Systems and 3D-Layer Printing

Stereolithography is the oldest and most widely used method for 3D printing or materials fabrication by additive manufacturing. Charles Hull (1986) coined the term in a patent issued in 1986 as a method for making solid objects by successively forming (or printing) thin layers of ultraviolet (UV) or photosensitive, cured polymer resins in a liquid photopolymer resin vat. Figure 1 shows the principle features for this additive manufacturing methodology. A focused UV laser beam is selectively and systematically scanned across the surface of a photosensitive liquid polymer resin representing a cross section of a modeled part to cure and solidify (by polymer cross-linking) this layer section. The beam is scanned by a CAD model

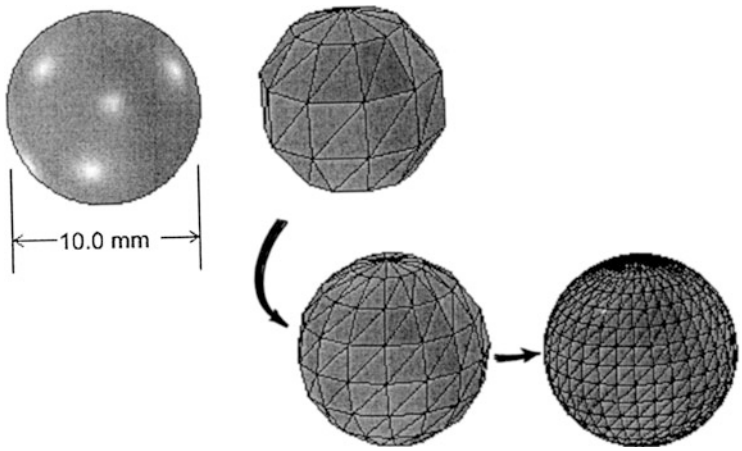


Fig. 2 STL file creation using object-surface triangulation to achieve object resolution

embedded in a digital file which directs the scanning mirror system shown in Fig. 1. This process is repeated slice by slice for the CAD model which forms a part in a corresponding layer-by-layer 3D-building process. Between each layer building the cured (completed) layer is swept by a leveling or recoater blade which is moved across the surface to smooth it and remove excess, uncured resin. The build platform is correspondingly lowered by a distance equal to the layer thickness, and a subsequent layer of resin is then wiped across the surface, selectively cured by the rastered laser beam, and the process repeated until the sectioned model is fabricated as a 3D object. The fabricated product is then removed from the resin vat and further cured in a UV cabinet.

The slice or layer-section data embedded in the CAD file and operating software which directs the scanning mirror in Fig. 1 to selectively scan the focused laser beam over a corresponding resin surface area to be cured and solidified, originates as an *STL* (STereoLithography or Standard Tessellation Language) *file*. This file format is supported by a number of software packages and formalisms for rapid prototyping and a variety of computer-aided manufacturing (CAM) formats. The STL format specifies binary and ASCII representations, with the former being the more common file format (Chua et al. 2003).

An STL file is a triangular representation of a 3D object: a triangulated surface by the unit normal and vertices of the triangles using a 3D Cartesian coordinate system. Every edge is part of exactly two triangles and not self-intersecting. The surface of a 3D object is broken into a logical series of triangles as illustrated for a solid sphere in Fig. 2. Each triangle is uniquely defined by its normal and three points representing its vertices. The resolution of surface detail improves with triangle refinement which also correspondingly increases the STL file size, which for good object reproduction ranges from 1 to 5 MB (Chua et al. 2003). Triangulated models can be constructed

mathematically or from scanned data such as CT or MRI. In addition, reverse engineered, complex-shaped objects can be replicated by metrology instruments such as coordinate measuring systems, especially laser triangulation systems which reconstruct x, y, and z coordinates corresponding to the laser line points by combining information provided by the laser plane intersection with the surface to be measured (Wang et al. 2002).

Complex structures, especially those consisting of an interconnected network of ligaments such as isotropic, random, open-cellular foams can be generated using mathematically based software such as *Surface Evolver* (Bakke 1992) illustrated in Fig. 3. Figure 3a shows an isotropic random foam generated by Surface Evolver (Bakke 1992) while Fig. 3b shows a group of cells extracted from this random foam. Figure 3c shows a skeletal drawing of the cell edges representing the foam section in Fig. 3b. Converted to an STL file, the structure can be sliced into layers used in the stereolithographic fabrication to build the model in Fig. 3a, b layer-by-layer as implicit in Fig. 3d.

Structure generation, especially complex open-cellular foam structures implicit in Fig. 3, can also be developed from micro-CT scans of actual foam materials, such as aluminum foams from which a build unit or structure generator can be identified and used to expand the stereolithographic product to specific cell sizes or characteristic porosities. Software packages, such as *Materialise/Magics* (http://www.materialise.com), utilize geometrically defined structure generators composed of a so-called unit cell or lattice structure unit similar in concept to a crystal lattice unit cell. Figure 4 illustrates a series of these unit cell build elements and their associated 3D models while Fig. 4a shows a stereolithography-fabricated polymer product based on the dode-thin model shown in Fig. 4, while Fig. 5b illustrates a more complex 3D-fabricated stereolithography product.

In the stereolithographic fabrication of many products, supporting structures are required which attach the product to the elevator platform and hold cross sections in place and resist recoater blade pressure during layer sweeping. These supports are generated automatically during CAD-file development and must be manually removed from the finished 3D product. Most stereolithography systems can produce parts with a maximum size of roughly 50 × 50 × 60 cm, but larger units can build on a platform allowing parts as large as 225 × 75 × 85 cm. The laser beam diameter is normally focused to ~150–300 μm, with scanning speeds of ~3 m/s. The build platform can be moved vertically by as little as ~2 μm. Common photopolymers are ABS-like resins and other polymer – like liquid UV – curable resins. They are sometimes free radical systems based on acrylate and methacrylate monomers. These cure rapidly and can be functionally modified to achieve materials with a variety of properties. Infrared (IR), thermo-initiated process fabrication using thermosensitive resins is also utilized in some stereolithographic, 3D-printing systems.

It should also be apparent that complex, internal structure products, such as closed-cell foams (Fig. 6a) cannot be fabricated by stereolithography as well as some of the other additive manufacturing or 3D printing processes (such as SLS)

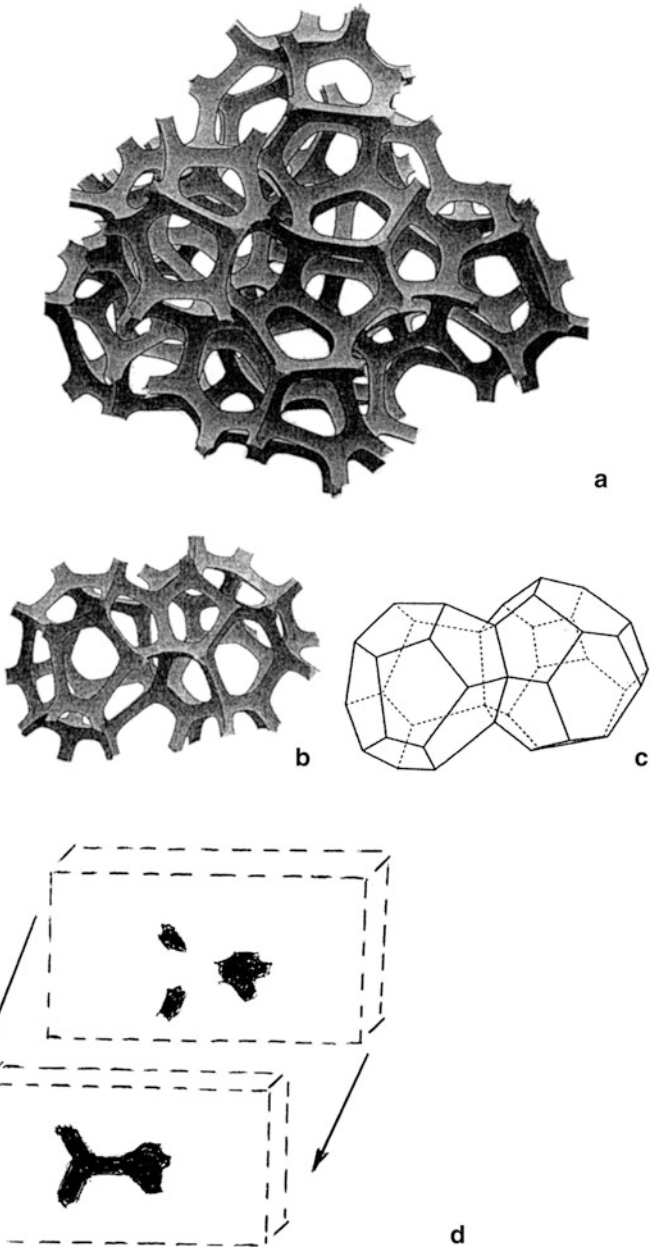


Fig. 3 Complex foam model generation and stereolithographic component replication. (**a**) Isotropic random foam image constructed by *Surface Evolver* (Bakke 1992). (**b**) Cell section extracted from (**a**). (**c**) Irregular, polyhedral cell line model of (**b**). (**d**) Layer building of foam section by stereolithography. Advancing slices are indicated by *dotted* layer sections

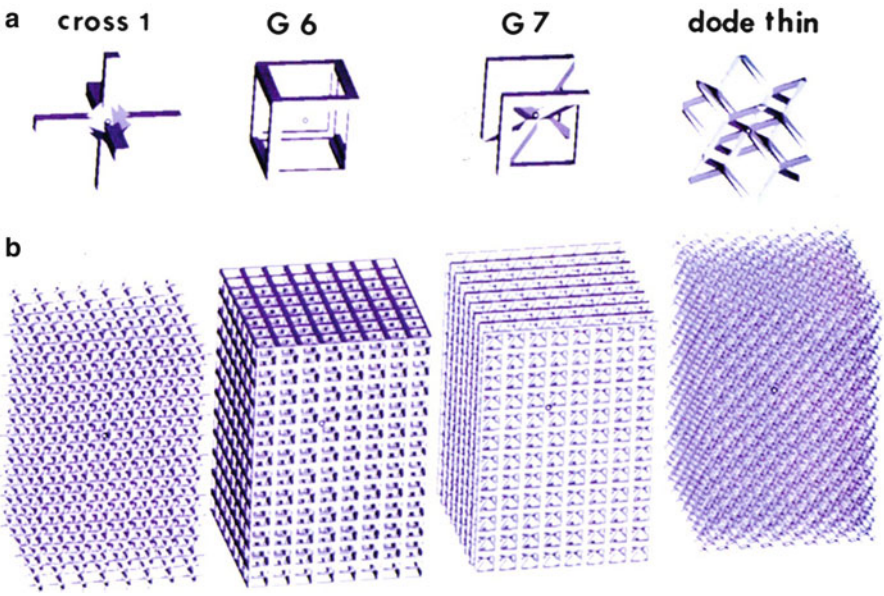


Fig. 4 Materialise™ software elements (**a**) and corresponding 3D models (**b**)

because these structures will retain the uncured resin or unsolidified powders. Open-cellular foams (Fig. 6b) and related, interconnected, internal structures can allow uncured resin and unsolidified powder to be drained or recovered.

There are numerous applications of stereolithography in a wide range of product prototyping and commercial product fabrication. There are numerous biological applications including the use of MRI or CT imaging to create patient-specific, preoperative models for surgical procedures to reduce the time of the operation thereafter as well as the risks (Sarment et al. 2003; Mankovich et al. 1994). The fabrication of a variety of tissue engineering scaffolds having requisite porosity and porous structures as implicit in Fig. 5a is also well established (Melchels et al. 2010).

Fused Deposition Modeling

While fused deposition modeling (FDM) is a 3D-printing, rapid prototyping technology, it differs from stereolithography in layer building which is similar in some respects to continuous inkjet printing since it uses dispenser heads to extrude liquid, thermopolymers into continuous, aligned tubes which fuse together to form a layer or layer portion representing an STL file model slice. This CAD-model printing system is similar to the stereolithography STL file system except the software drives the polymer extrusion and its deposition from x-y translating, dispensing heads: one normally depositing the product material and the other the corresponding supporting material, which is often water soluble so that it can be

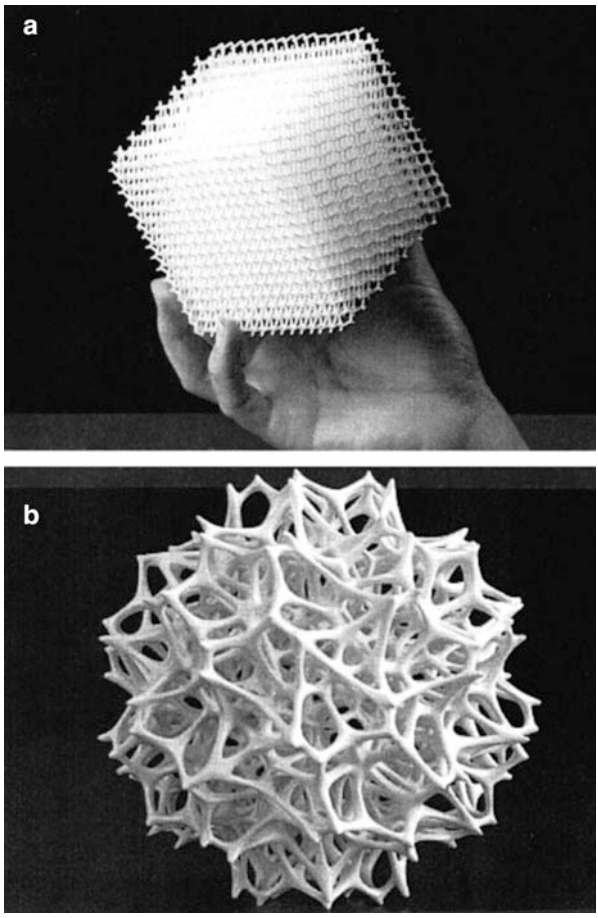


Fig. 5 Stereolithographic 3D polymer product examples. (**a**) Reticulated mesh using the "dode-thin" model in Fig. 4b. Courtesy of Milwaukee School of Engineering. (**b**) Complex open-cellular structure having nonlinear ligaments (Courtesy of Rene Schmidt)

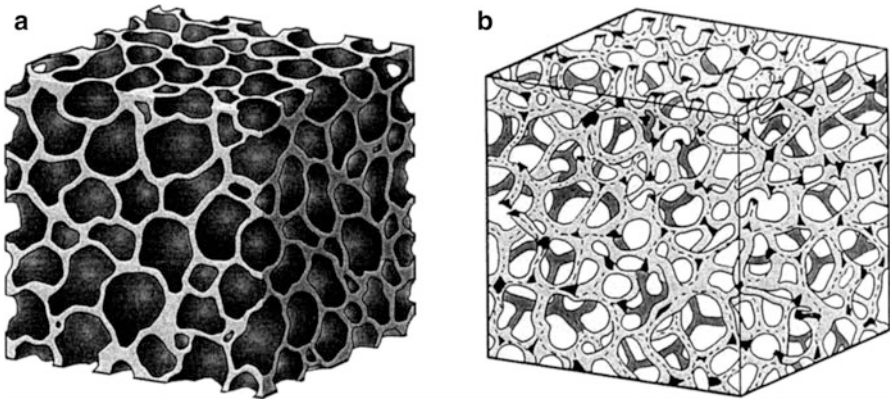


Fig. 6 Foam models. (**a**) Closed cell. (**b**) Open cell

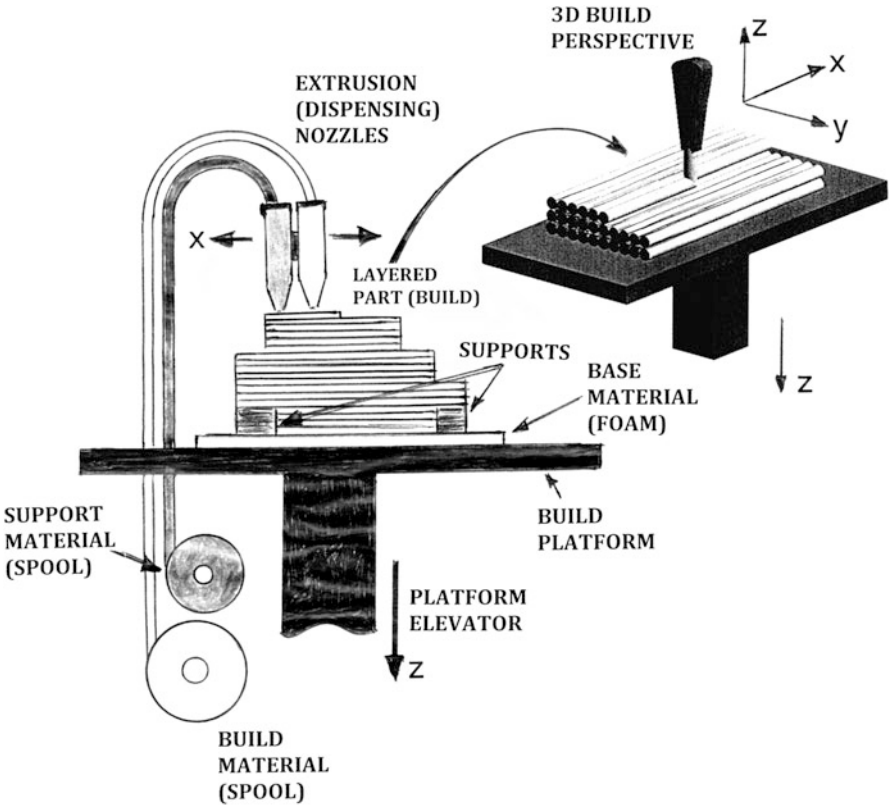


Fig. 7 Schematic view showing the fused deposition modeling (FDM) process for 3D printing. The build material can be in spool form or pumped liquid. The build and support materials are heated in the dispensing heads to some predetermined viscosity or flowability

easily removed by rinsing or submersion in a water bath following 3D product fabrication. Figure 7 illustrates this SFF system schematically. Commonly used thermoplastic elastomers include polyethylene, polypropylene, and especially polycarbonate and polycarbonate resins and acrylonitrile butadiene styrene (ABS) which are illustrated in Fig. 8. ABS is a copolymer made by polymerizing styrene and acrylonitrile in the presence of polybutadiene. This results in a long chain polybutadiene crisscrossed with shorter poly(styrene-co-acrylonitrile) chains. ABS is stronger, more rigid, and more dimensionally stable than polystyrene. It might be useful at this point to provide a very brief review of polymer structure, at least as it differentiates a copolymer such as ABS from other common polymer structure forms. Polyethylene is used as a primary reference as shown in Fig. 9.

FDM (Russell and Cohn 2012) was developed by Scott Crump in the late 1980s who founded Stratasys, one of the first manufacturers of 3D printing machines based on FDM. The term 3D printing was a trademark of Stratasys, Inc. which they allowed to enter the public domain and become a generic industry term in 1999.

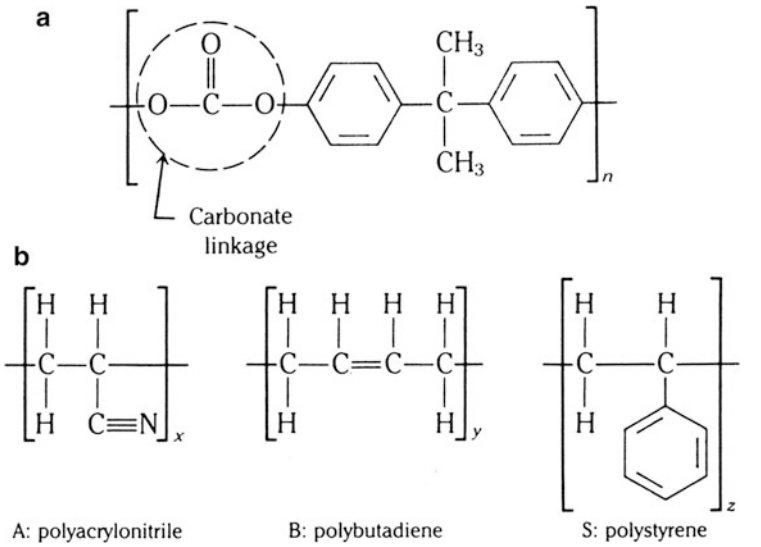


Fig. 8 Chemical structural (monomer) units representing common FDM thermopolymer build materials. (**a**) Polycarbonate (~270 °C melting point). (**b**) ABS (which melts at 105 °C but softens between 50 °C and 80 °C)

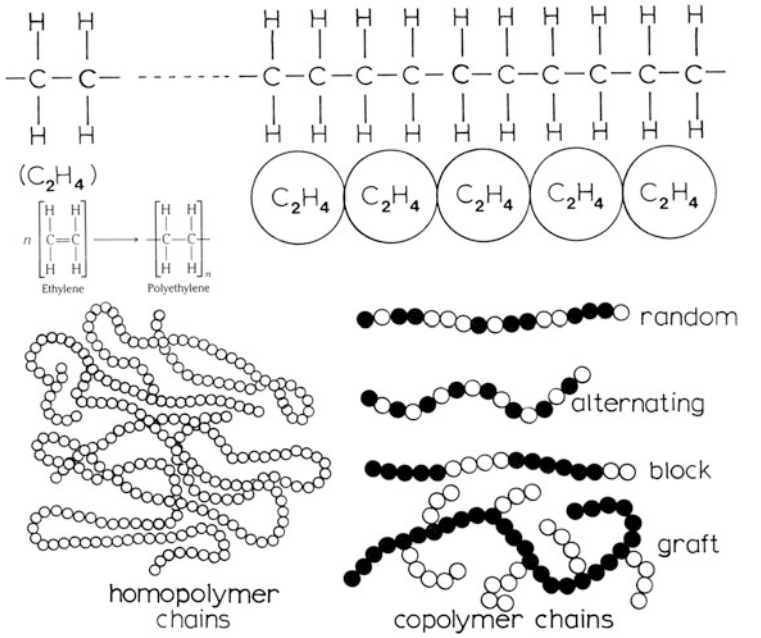


Fig. 9 Common polymer structure forms. Copolymer chains also characterize cross-linking of either homopolymer chains or other polymer chain types shown

As illustrated in Fig. 7, the thermoplastic build and support materials in spool forms are fed into heated print heads or dispensing heads with nozzle designs to optimize the extrusion of the liquid or semiliquid thermopolymers, whose viscosities can be adjusted somewhat to facilitate fusion of the extruded and deposited polymer. Depending upon the build and support material compositions, nozzle erosion can occur and nozzle replacement must be considered. The build table is usually cooled to control the product layer development, and the table is moved down to accommodate the deposition of the next layer. Many FDM systems can print layers as small as 160 μm, and normal product layer thickness varies from ~200 to 350 μm, with extruded thermoplastic widths between ~250 and 1,000 μm (1 mm).

The support material (Fig. 7) is a different color than the product building material and does not fuse or adhere to it. This allows the support material to be easily identified and removed. It should also be apparent that unlike stereolithography (Fig. 1), FDM can build complex, closed-cell or related internal structures (Fig. 6a). In addition, multihead or multi-jet dispensing systems can allow for complex or even embedded structure fabrication, or the more rapid building of complex models. Many, commonly available 3D printing systems are FDM-based SFF technologies.

Because the thermopolymer build material can be developed as a spooled precursor rod or fiber of various diameters which is melted in the dispensing or jet head prior to extrusion (Fig. 7), it is possible to create polymer matrix/composite materials by suspending ceramic particles, including nanoparticles, within the thermopolymer matrix. These can also include nano- or microrods, even carbon nanotubes, which may orient themselves somewhat in the extruding and layer-dispensed material to build layer-specific composite structures having superior strength or tribological properties. By loading the polymer matrix with a high fraction of ceramic particles, a 3D-printed product can be heated at high temperature to evaporate the polymer binder and sinter the product at an appropriately elevated temperature. Of course this product would incur necessary shrinkage which could be incorporated into the design and fabrication of the final product. 3D printed products are also sometimes referred to as 4D-printed products if they change their shape or self-assemble or reassemble with time after 3D fabrication. Time is always the fourth dimension.

It is possible to create a hybrid stereolithography-fused deposition modeling configuration by dispensing less viscous, photocurable polymer resins from the printing head illustrated in Fig. 7 and simultaneously curing the dispensed photopolymer stream. The photocuring (UV) laser is mounted alongside the dispensing head and is focused so that it cures the dispensed polymer stream in an optimal manner. This process has the advantages of FDM but allows for variances in the dispensing head to create somewhat more refined (higher resolution) products. This process can also create complex internal structures like FDM without trapping any excess resin which is a limitation of conventional stereolithography processing in a photocurable resin bath as illustrated in Fig. 1. It can also be noted in the FDM

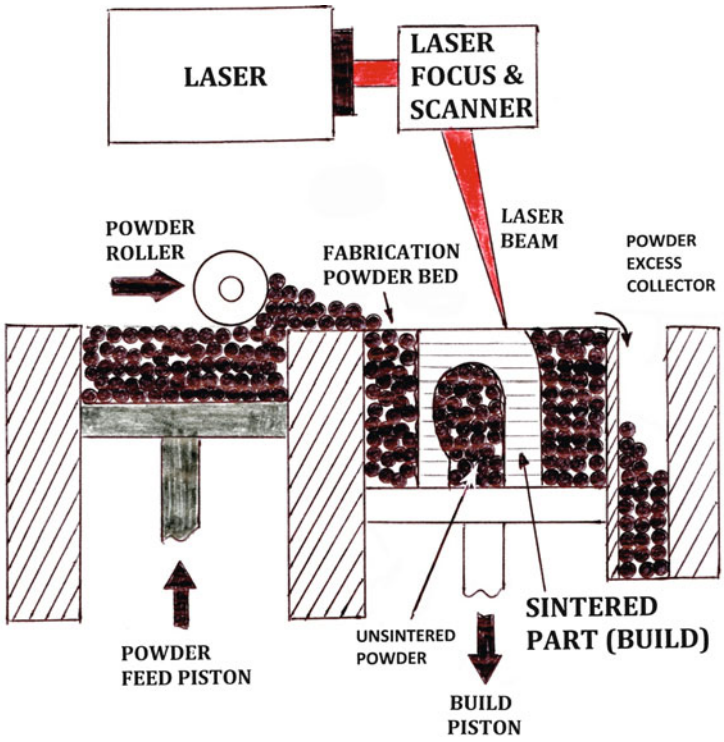


Fig. 10 Selective laser sintering (SLS) process schematic. Note that excess powder is rolled to a collector at *right*. Unsintered powder in the *center* of the build is removed after fabrication

process (Fig. 7) that multi-material deposition and fabrication is possible using different dispensing heads. For example, the dispensed material can be side by side or in layer configurations if desired.

Selective Laser Sintering

As illustrated schematically in Fig. 10, selective laser sintering (SLS) utilizes a powder bed of (thermo) polymeric material to create STL file slices of a CAD model which are selectively sintered using an appropriate laser-directed beam. While the process illustrated in Fig. 10 is generically referred to as selective laser sintering, the laser beam energy can either partially or fully melt the polymer particle powder layer (∼100–300 μm thick) or create a regime often referred to as liquid-phase sintering. While partial melting at the particle surface can allow sufficient fusion to produce a fully dense, layer-fabricated monolithic component having no adverse microstructural consequences, this is not the case for metal powders which, if not completely melted, contain a microstructural mixture (or a composite microstructure) with the unmelted particle core having essentially the

precursor particle microstructure and the melt region surrounding (and fusing) particles in the layer bed having a quite different microstructure. Even in the case of fully melted metal powder particles, the microstructure can be variously manipulated by layer cooling and the associated layer-by-layer melt phenomena which are much different from conventional melt phenomena associated with directional solidification and the corresponding solid/melt interfacial regime. Consequently, metal powder fabrication by either laser melting or electron beam melting constitutes a rather unique additive manufacturing process which will be discussed categorically in a following section.

In most SLS applications, a high-power CO_2 laser (operating at 10.6 μm wavelength) is employed. While the efficiency of SLS monolithic part building depends on the peak laser power in contrast to the powder melting point, rather than the dwell time of the laser beam on a layer portion, pulsed lasers are normally used. YAG ($Y_3AL_5O_{12}$) lasers (operating at a wavelength of 1.06 μm) and Na:YAG (neodymium-doped YAG) lasers which can operate down to 355 nm wavelength can also be used where lower power requirements can achieve appropriate liquid-phase sintering of the powder bed. It might be noted from Planck's relationship that smaller wavelengths impose characteristically higher energy (► Eq. 72 of chapter "Electromagnetic Fundamentals"). ABS polymer powders, for example, can be liquid-phase sintered or melted at temperatures as low as 80 °C (Fig. 8b).

SLS, like other additive manufacturing technologies, was developed in the mid-1980s by Carl Deckard and Joe Beaman at the University of Texas at Austin. As illustrated in Fig. 10, powder is rolled onto the build platform in layer or STL-slice thickness as noted. The directed laser beam selectively sinters or melts the layer which is also melted or sintered to the previous layer, and the build platform then drops down the thickness of the processed layer and a new layer rolled from the powder feed chamber to the left in Fig. 10. Since the unprocessed powder supports the building component, there is normally no need for specific support structures to be fabricated. However, like stereolithography processes illustrated in Fig. 11, complex structures cannot be created unless there is a continuous or open-cell structure which permits the removal of unprocessed powder.

In normal SLS processing, the powder bed is heated to accelerate and optimize the sintering process usually by heating the workable area to just below the melting point of the powdered polymer. Of course this imposes unusual requirements for metals as implied above, and especially ceramic powders, most of which melt at temperatures significantly above metal powder melting points. Ceramic products, however, can be produced by another additive manufacturing variance where the powder is rolled from a powder source as in Fig. 10, but instead of a laser beam selectively sintering the rolled layers, a dispensing head places a binder on the rolled ceramic powder layer as directed by the STL file slice model. This bound, additively built ceramic component is then removed from the build chamber and placed in a high-temperature furnace to evaporate the binder and allow the ceramic powder to sinter. This often produces considerable shrinkage and residual porosity, but more complex structures can be fabricated in this process than most other

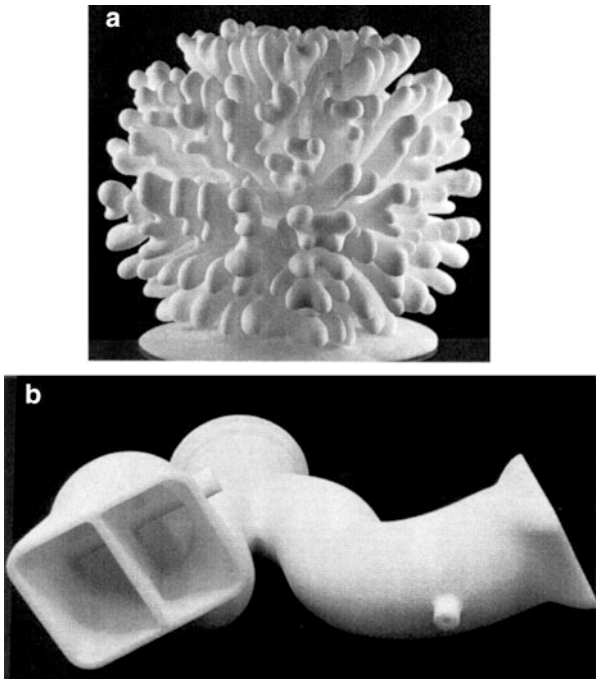


Fig. 11 Examples of SLS fabricated configurations and products. (**a**) Complex geometry component. (**b**) Commodity product; multiple part production example (Google images)

conventional processes. In a modification of this process, ceramic-binder mixtures can be extruded or dispensed as in the FDM process to form a so-called green product which is then furnace sintered as indicated above. Earlier uses of ceramic-binder processes, especially ceramic-binder tapes, were often referred to as laminated object manufacturing as a category of additive (layer) manufacturing. These could be laid down as laminates and trimmed by the laser as directed by the CAD program. Binders are often methacrylic acid derivatives. Other modifications of these technologies have involved mixing ceramic powders (such as Al_2O_3) with photocurable polymer resins and processed by stereolithography (Fig. 1), whereafter the fabricated product is sintered at high temperature simultaneously with the binder removal. Correspondingly, polymer-ceramic composite products can also be fabricated directly as indicated previously, where the ceramic, even in fiber form, strengthens the cored polymer matrix.

Although SLS can be either a limited-run (even prototypical) or a high-volume product-manufacturing technology, it is capable of producing any product geometry which can be embedded in an STL file, CAD model like stereolithography, or FDM. Consequently, complex open-cell (Fig. 6a) components such as those shown in Fig. 5 are readily fabricated by SLS, as well as a wide range of other commodity products and components as illustrated typically in examples shown in Fig. 11. SLS machines are manufactured worldwide and are used in a wide variety of additive manufacturing venues along with FDM systems, which are being manufactured as personal 3D printers available in a very general market. These

are being used in every conceivable product-manufacturing arena from complex chocolate products to intricate, movable plastic commodity items, that is, products with as-fabricated movable and interconnected parts such as gears, etc. This has been referred to as next generation manufacturing or the "third industrial resolution" (Whadcock 2012).

References

Bakke KA (1992) The surface evolver. Exp Math 1:141–165

Bartolo PJ (2011) Stereolithography: materials, processes and applications. Springer, New York

Chua CK, Leong KF, Lim CS (2003) Chapter 6, Rapid prototyping formats. In: Rapid prototyping: principles and applications. World Publishing, Singapore, p 237

Gibson I, Rosen DW, Stucker B (2010) Additive manufacturing technologies: rapid prototyping to direct digital manufacturing. Springer, New York

Hull CW (1986) Apparatus for production of three-dimensional objects by stereolithography. US Patent 4,575,330

Jacobs PF (1992) Rapid prototyping and manufacturing. SME, Dearborn

Mankovich NJ, Samson D, Pratt W, Lew D, Breumer J (1994) Surgical planning using 3-dimensional imaging and computer modeling. Otolaryngol Clin North Am 27(5):875–889

Melchels FPW, Fejpen J, Grijpma DW (2010) A review on stereolithography and its applications in biomedical engineering. Biomaterials 31:6121–6130

Russell J, Cohn R (2012) Fused deposition modeling. Wikipedia, Bookvika Publishing (on demand publishing format)

Sarment DP, Sukovic P, Clinthorne N (2003) Accuracy of implant placement with a stereolithographic surgical guide. Int J Oral Maxillofac Implants 18(4):571–577

Wang G, Zheng B, Li X, Houkes Z, Regtien PPL (2002) Modeling and calibration of the laser beam-scanning triangulation measurement system. Robot Auton Syst 40:276–277

Whadcock I (2012 April 21) A special report: a third industrial revolution. The Economist :3–21

Digital Fabrication with Embedded Electronics

Contents

Abstract

The fabrication of integrated circuits by lithography technologies is a limited dimension 3D printing concept which can be expanded by building larger dimension structures in which mechanical-electromechanical and related device structures can be embedded. These involve hybrid, interconnected, layered, functional electronics along with power sources and other device structures fabricated by either some form of 3D printing or other novel placement.

Introduction

In the same way components are added to the surfaces of electronic boards and systems, future processes will incorporate an even greater variety of electronic, mechanical-electromechanical systems (*MEMS*), optical systems and devices, and related devices as well as power sources into 3D printed or other layer-integrated, 3D products. These technologies will involve not only a variety of materials dispensing strategies but also a similar variety of materials having specific properties and performance features. In this short chapter a few contemporary examples are provided to illustrate these concepts.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_38

Narrative

While numerous examples of rapid prototyping or solid freeform fabrication and their hybrids have been presented previously in chapters "▶ Bioprinting and Biofabrication of Organ" and "▶ Rapid Prototyping Technologies: Solid Freeform Fabrication," it is certainly apparent that other important variances of digital fabrication or additive manufacturing are emerging. This is especially true for many commercial, aerospace, and related military applications where miniaturization and functionality are key factors. Truly 3D products, which include multilevel and interconnected electronic components along with other functionalities fabricated digitally (directly from digital design (CAD) files) by processes generally involving 3D printing, are becoming more in demand as traditionally packaged electronic device components, including integrated circuits and surface-mount devices, are reaching a point where the package size is too large to achieve the desired device density, on the available circuit board surface. A better approach or solution which is evolving includes multitask 3D printing systems or hybrid additive manufacturing systems able to create layered, functional electronics interconnected with embedded, unpacked components, including bare ICs (or bare die semiconductor devices), sensors, or even microelectromechanical systems (MEMS) components. This affords the integration of analogue (MEMS) and digital (electronic) application-specific integrated circuits (ASICs). Here the interconnects can involve planar, layered printed conductors or vertical (via) interconnections.

It can be appreciated on examining the 3D printing issues and fundamentals implicit in ▶ Figs. 1, ▶ 7, and ▶ 10 of chapter "Rapid Prototyping Technologies: Solid Freeform Fabrication" in retrospect, as well as drop-on-demand printing illustrated earlier in ▶ Fig. 6 of chapter "3D Printing: Printed Electronics", that these functions could be integrated into a complex, interchangeable, multifunctional, and multi-material additive manufacturing regime. This manufacturing regime is capable of layer-by-layer building of complex, physical structures including electronic device components, optical elements, MEMS arrays, and the integration of IC dies (or chips) and other MEMS (analogue) dies into the build process using variations of pick-and-place tools to mount these devices within the building monolith. ▶ Figure 1b of chapter "3D Printing: Printed Electronics" illustrates the placement of surface-mount devices on a flexible circuit, or circuit printed on a flexible polymer substrate. More delicate device transfer can also utilize laser-based device transfer processes referred to as "lace and place," a contactless technique that allows the transfer of very small, thin devices not readily amenable to pick-and-place tools (Pique and Chrisey 2012). In effect SFF systems can integrate many different materials and multiple building processes to fabricate complex, functional (or multifunctional), and integrated electronic and electromechanical devices in a monolithic, 3D product (Lopes et al. 2012). Figure 1 illustrates a commercially available SFF system with 4 dispensing heads, a CO_2 or YAG laser, and a pick-and-place system, all with precise 3D motion and material control. The dispensing heads can accommodate viscosities over a range of 6 orders of magnitude.

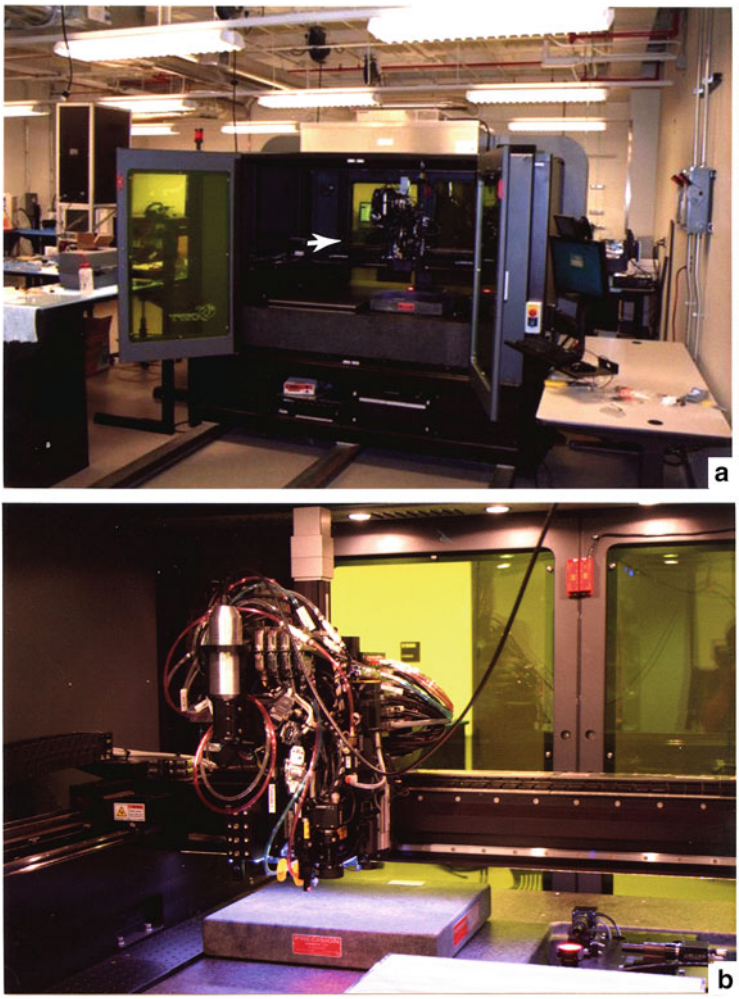


Fig. 1 Commercial (nScript, Inc.) multi-head, 3D printing machine. The arrow in the overall view in (**a**) shows the multi-head, multifunction SFF build components illustrated in the enlarged view in (**b**) (Courtesy of W.M. Keck Center for 3D Innovation, University of Texas at El Paso)

Figure 2 illustrates a 3D fabricated component with embedded electronics. Figure 2 is a 3D printed gaming die. When the die is rolled, an embedded microcontroller and accelerometer identify the upper surface and illuminate the corresponding LED display as shown in Fig. 2b. Figure 2a illustrates the CAD model for Fig. 2b, while Fig. 2c illustrates a model view showing the interior structure which contains the 3-axis MEMS accelerometer and a lithium polymer rechargeable battery shown in the component view for the actual fabricated component in Fig. 2d, while Fig. 2e shows the fabricated outer die and inner die

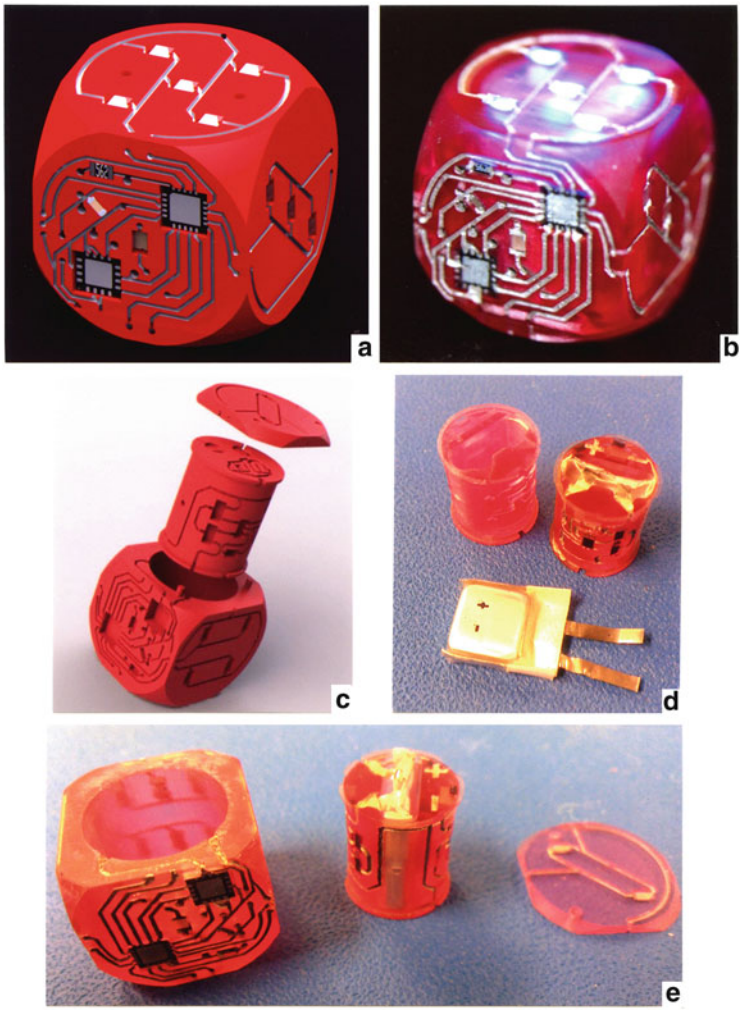


Fig. 2 Integrated 3D printed gaming die. (**a**) CAD model. (**b**) Fabricated die corresponding to (**a**) with LED illustrated "S" face. (**c**) Exploded CAD model showing die interior. (**d**) Interior fabrication showing battery. (**e**) Prefabricated components representing integrated 3D die in (**b**) (Courtesy of D. Muse and E. MacDonald (MacDonald et al. 2013), W.M. Keck Center for 3D Innovation, University of Texas at El Paso)

components, including the battery. Other major components include an 8-bit microcontroller, N-channel MOSFET transistors, a lithium polymer protection/safety IC, capacitor and resistor chips, and the LEDs in the die faces as illustrated in Fig. 2b. The substrate or die material was a stereolithography photopolymer (ProtoTherm 12120 made by DSM Somos), while the conductive ink interconnects were printed using an Ercon E1660 silver flake-based ink.

While the 3D-embedded electronic component shown in Fig. 2 represents a hybrid prototype assembled by separate but integrated 3D printing technologies, it is becoming possible to fabricate such intricate, 3D embedded electronic components in a single additive manufacturing operation using commercial systems illustrated conceptually in Fig. 1.

References

Lopes AJ, MacDonald E, Wicker RB (2012) Integrating stereolithography and direct print technologies for 3D structural electronics fabrication. Rapid Prototyp J 18(92):129–143

MacDonald E, Muse DW, Keck WM (2013) Center for 3D innovation. University of Texas at El Paso, Private communication

Pique A, Chrisey DB (eds) (2012) Direct write technologies for rapid prototyping applications: sensors, electronics and integrated power sources. Academic, San Diego

Novel Structure Printing

Contents

Abstract

Novel structure printing is illustrated through the fabrication of negative Poisson's ratio or auxetic structures and negative index of retraction structures or cloaking structures and stealth configurations.

Introduction

Before discussing the wide range of metal and alloy melting and related additive manufacturing strategies using metal powders, it might be interesting to examine a few novel, if not unusual, 3D printing or additively fabricated structures. The structural issues here will be novel geometrical structures conducive to fabrication by 3D printing technologies.

Negative Poisson's Ratio (Auxetic) Structures

In the first of these novel structures, negative Poisson's ratio structures introduced briefly in chapter "► Biomimetics and Biologically Inspired Materials" and illustrated specifically in ► Figs. 6 and ► 7 of chapter "Biomimetics and Biologically Inspired Materials" will be revisited. ► Figure 7b of chapter "Strategies for Bone Replacement and Tissue Augmentation" shows novel implants or implant

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_39

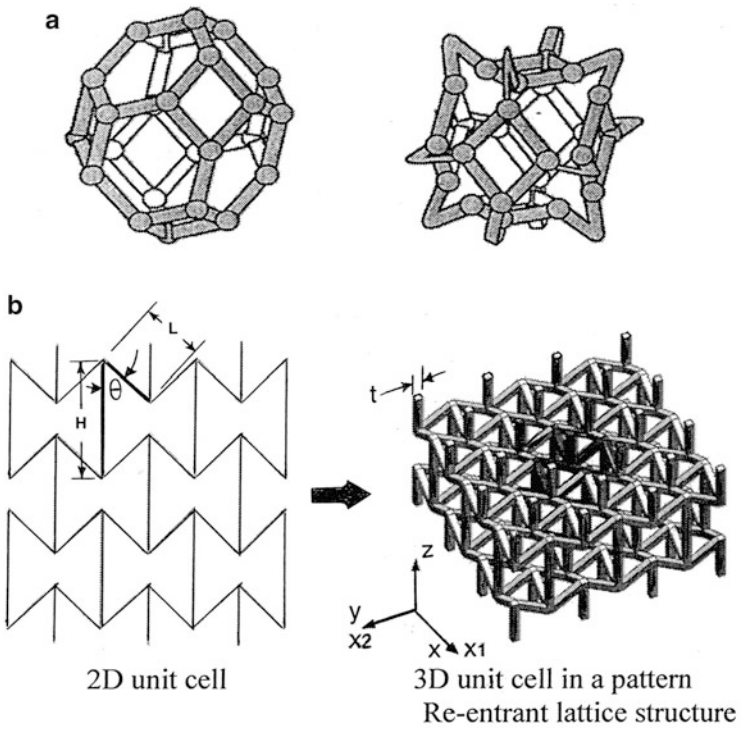


Fig. 1 Comparison of regular truncated cuboctahedron (*left*) and reentrant (unit) forms (*right*) (**a**), with 2D and 3D reentrant unit cells and 3D reentrant lattice structure (**b**). The unit cell element is shown by bold lines in (**b**). (**a**) is adapted from (Prawoto 2012), while (**b**) is adapted from Yang et al. (2011)

geometries (auxetic geometries) which render a unique advantage for both polymer and metallic materials as a consequence of the resultant negative Poisson's ratio performance. By defining a basic unit cell to describe simple auxetic geometries which can be modeled in CAD as implicit in ▶ Figs. 3 and ▶ 4 of chapter "Rapid Prototyping Technologies: Solid Freedom Fabrication" for conventional or regular structures, negative Poisson's ratio structures can be built by various 3D printing technologies using models embedded in STL files. Figure 1 illustrates the basic principle for these structures. Figure 1a compares a conventional truncated cuboctahedron (3D) unit compared to an auxetic or reentrant unit. *Reentrant* refers to an angle in a polygon greater than 180° (a negative angle) which in effect points inward. This concept is illustrated more specifically in the 2D-to-3D unit cell (reentrant lattice structure) comparison shown in Fig. 1b. Here the force components can be shown to be given by (Yang et al. 2011)

$$F = \frac{\sigma}{2}\,(2\,L\sin\theta)^2 = 2\,\sigma\,L^2\,(\sin\theta)^2 \qquad (1)$$

The 3D reentrant structure in Fig. 1b is orthotropic with two identical directions (x1 = $x2$) representing the x-y plane and Z representing a third direction which can be considered to be the layer normal or build direction in any form of stereolithography, or 3D printing technology. From standard beam theory as implicit in ▶ Fig. 88 of chapter "Chemical Forces: Nanoparticles" for struts having a cross-section thickness, t, the Poisson's ratio becomes (Yang et al. 2011):

$$\nu = -\frac{L^2 \cos\theta(\,\alpha - \cos\theta)}{2\alpha t^2 + L^2 \sin^2\theta} \tag{2}$$

where α = H/L from Fig. 1b. The corresponding structural modulus (or Young's modulus) for the reentrant structure in Fig. 1b becomes

$$\mathrm{E}' = \frac{\sigma}{\varepsilon'} = \frac{2(\alpha - \cos\theta)}{4\alpha t^2 + L^2 \sin^2\theta} \left(\frac{t^2}{L\sin\theta}\right)^2 \mathrm{E} = \frac{\mathrm{FL}\sin\theta}{8} \tag{3}$$

where E and σ are the modulus and yield stress for the bulk (solid) material, respectively. Consequently, in reexamining ▶ Eq. 7 in chapter "Implant Materials and Structures", the specific stiffness versus specific density for an auxetic or reentrant structure can be written in the form

$$\frac{E_r}{E_s} = 0.05\zeta \left(\frac{\rho_r}{\rho_s}\right)^2, \tag{4}$$

where the subscript r denotes the reentrant structure and s denotes the solid, fully dense structure; ζ is a geometrical function depending upon the reentrant form (Prawoto 2012). It can have value between 10 and 20.

Greaves et al. (2011) have recently reviewed some engineering applications of auxetic (reentrant) materials structures, while Yang et al. (2011) have recently fabricated a 3D reentrant, open-cellular structure using the design unit cell and 3D lattice structure model shown in Fig. 1b. The structure fabricated by Yang et al. (2011) employed electron beam melting of Ti-6Al-4V powder, to be described in the next section in more detail. This structure also exhibited interesting energy adsorption applications where weight may be a limiting criterion. In connection with this phenomenon it might be noted that the hardness (or effective hardness) increases for auxetic structures as illustrated previously in ▶ Fig. 18 of chapter "Biomimetics and Biologically Inspired Materials. Prawoto (2012) has also recently reviewed computational mechanics aspects of auxetic structures, including this feature. Additive manufacturing offers a unique ability to fabricate an array of polymer and metal reentrant structures, although the size, and correspondingly the function, is limited by the build volume available in commercial system.

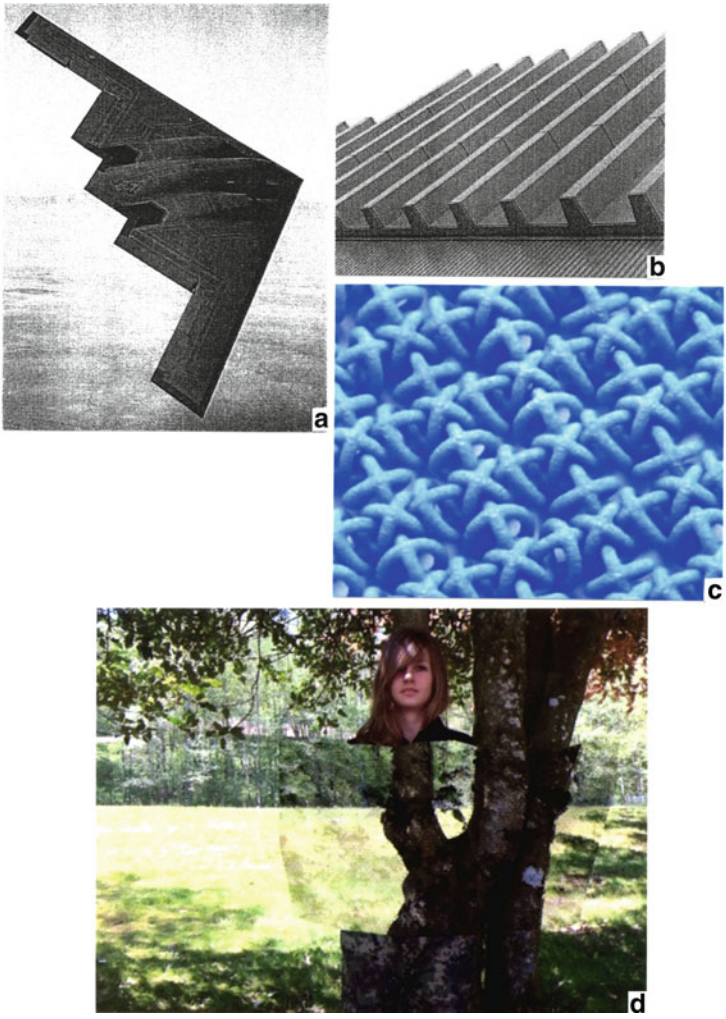


Fig. 2 Stealth structures and examples. (**a**) Northrop-Grumman B-2 Spirit radar stealth bomber. (**b**) Stealthy airport buildings structures for radio wave diffraction (Courtesy of H. Grouse, Airbus S.A.S.). (**c**) 3D printed fabric (Google Image). (**d**) Invisibility fabric or cloak (Courtesy of Hyperstealth Biotechnology, British Columbia, Canada)

Negative Refractive Index (Cloaking) Structures

The issue of build volume or build capacity in 3D printing technologies is also of some consequence in considering the 3D printing of negative refractive index structures, or more generally cloaking or variously stealth structures and geometries, or metamaterials as previously discussed in Part III. In the case of more

familiar stealth technologies which render structures ideally invisible to radar, infrared, and other detection wavelengths shapes are used to redirect these electromagnetic waves. In these applications, dielectric composites are more transparent to radar, while conductive materials such as metals or carbon fiber composites are reflective. Flat, facet surfaces such as those characteristic of the B-2 bomber shown in Fig. 2a concentrate scattered radar energy primarily into a forward scatter direction and away from the source. It might be recalled that over-the-horizon radar is characterized by a wavelength ranging from 10 to 100 m (corresponding to frequencies ranging from 3 to 30 MHz). Similarly, Fig. 2b illustrates a section of what is referred to as "stealthy airport buildings technology" developed by Airbus. In this concept, aluminum panels with appropriate shapes and dimensions diffract radio waves to stop them from reaching airport runways. It might be noted that Fig. 2b emulates the diffraction grating features for insect structures as noted in ▶ Fig. 8b, c of chapter "Rapid Prototyping Technologies: Solid Freeform Fabrication", but for spacings roughly $>10^8$ greater (meter sizes versus hundreds of nanometers).

Urzhumov et al. (2013) have recently described the stereolithographic polymer-based fabrication of a microwave X-band (~10 GHz) cloaking device. The device is a shell of an air-dielectric composite having an array of holes of various shapes and sizes to deflect microwave radiation. The disclike cloak has a circular opening at its center where an opaque object can be placed. When microwave beams are aimed at the object through the size of the disk, the object was essentially invisible because of reradiation back into free space on the shadow side of the cloak.

This concept might be extended to complex shapes or shape arrays through fabric printing illustrated in Fig. 2c. Such fabrics could ideally be composed of metal and dielectric fibers or fiber arrays to create a requisite metamaterial for negative refractive index behavior in the visible portion of the spectrum. Figure 2d shows such an "invisibility cloak" using a requisite fabric structure which has not been disclosed by the manufacturer, but which certainly could be reverse engineered and embedded in CAD to allow 3D printing.

References

Greaves G, Greer A, Lakes R, Rouxel T (2011) Poisson's ratio and modern materials. Nat Mater 10:823–837

Prawoto Y (2012) Seeing auxetic materials from the mechanics point of view: a structural review of negative Poisson's ratio. Comput Mater Sci 58:140–153

Urzhumov Y, Landy N, Driscoll T, Basov D, Smith DR (2013) Thin low-loss dielectric coatings for free-space cloaking. Optics Lett 38(10):1606–1608

Yang L, Harrysson O, West II H, Cormier D (2011) Design and characterization of orthotropic re-entrant auxetic structures made via EBM using Ti6Al4V and pure copper. In: Bourell D (Ed) Proceedings of the 21st solid freeform fabrication symposium. University of Texas, Austin, TX, pp 464–474

Laser and Electron Beam Melting Technologies

Contents

Abstract

Laser and electron beam melting technologies are presented in this chapter along with a range of examples of metal and alloy products fabricated by these technologies. Laser beam melting or selective laser melting (SLM) in contrast to electron beam melting (EBM) processes metal or alloy powders in an inert atmosphere as opposed to a vacuum. Complex melting and solidification of the layers invokes new directional solidification metallurgy.

Introduction

Although ▶ Fig. 10 in chapter "Rapid Prototyping Technologies: Solid Freeform Fabrication" provides a general view for the selective laser sintering (SLS) process as previously discussed primarily for polymer and other low-melting material powders, there are notable differences which need to be addressed for metal and alloy powder processing using either a laser or electron beam. One notable issue is that in sintering polymer powders, the partially melted particle perimeter defining the sinter regime is not altered in contrast to the unmelted powder particle center. However, for metals the unmelted particle center microstructure would normally differ, even significantly, from the melted perimeter. In addition, the thermokinetics (such as rate of cooling of the sintered surrounding zone) will also influence the microstructure, especially the residual, solid-state microstructure.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_40

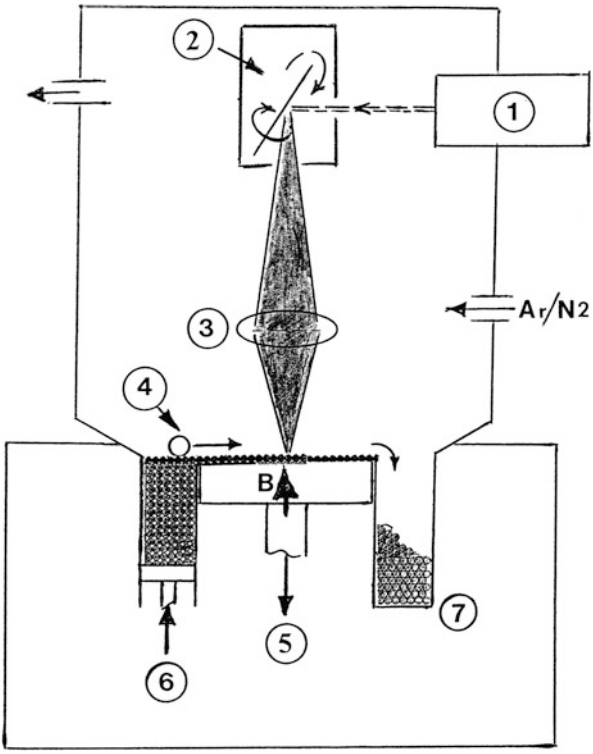


Fig. 1 Selective laser melting (SLM) schematic. Numbered components are described in the text. *B* denotes the part build direction on the build platform which moves down with each selectively melted layer and the rolling (*4*) of a new layer. Building is normally performed in a purified argon (*Ar*) or nitrogen (N_2) gas environment

Additive (layer) manufacturing using laser and electron beams as point source heating technologies in fact selectively fuses or melts the associated metal or alloy powder bed. This is also called powder bed fusion (PBF). In this process the particles in the affected layer are completely melted. In addition, succeeding layers are melted to the preceding layer creating repeated solid/liquid layered zones. This process certainly differs markedly from traditional solidification from a continuous melt as a consequence of an undercooling at the solid/liquid (melt) interface.

Selective laser melting (SLM) systems (also sometimes called Direct Metal Laser Sintering Systems) became commercially available in the 1990s, mostly originating from German manufacturers. These systems, as illustrated conceptually in ▶ Fig. 10 of chapter "Rapid Prototyping Technologies: Solid Freeform Fabrication," use mirror deflection systems to scan the laser beam with pre- or post-focusing optics. While the mechanical movements of the scanning mirror permit accurate laser beam scanning up to ~15 m/s scan rate, electromagnetic scanning in electron beam melting systems allows more than 2 orders of magnitude greater electron beam scan speeds with corresponding accuracy. Laser power can vary from 0.5 to 1 kW for YAG, Nd: YAG, or Yb-doped silica glass fiber lasers, either continuous wave or pulsed, with wavelengths varying from

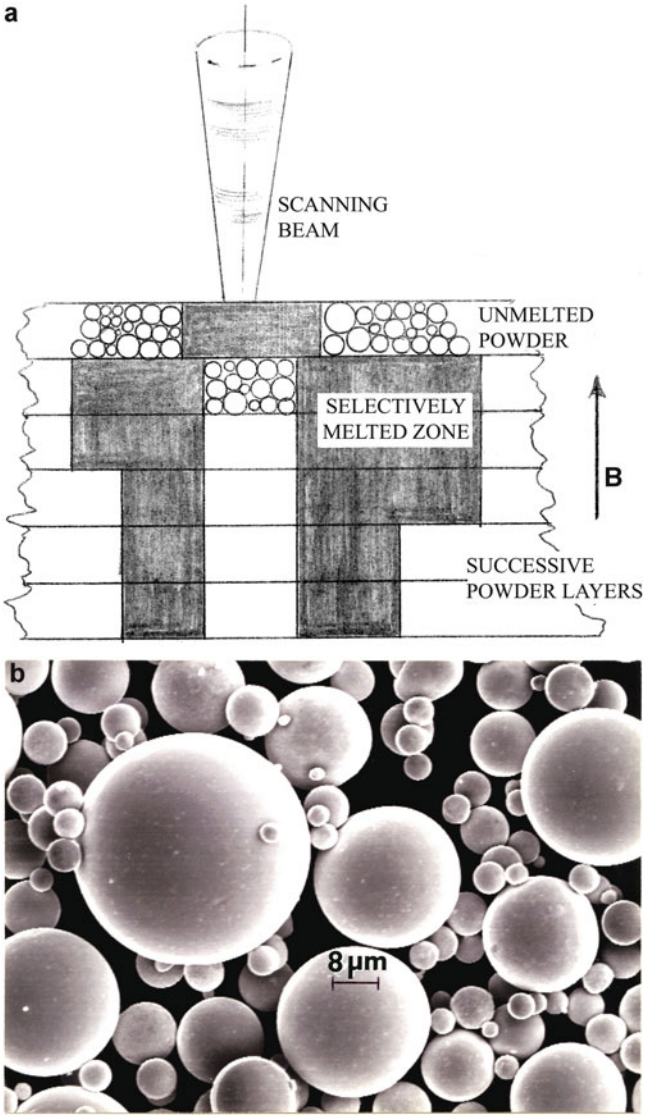


Fig. 2 Melt scan and selective powder layer melting to form complex part geometries from which unmelted powder can be removed (**a**). (**b**) shows an atomized, rapidly solidified copper powder example

$\sim$0.3 to $>$1 μm. CO_2 gas lasers have also been employed with energies $>$6 kW at a wavelength of 10.6 μm.

Both laser and electron beam melting technologies have their origins in welding technologies. Electron beam welding systems were originally manufactured in the

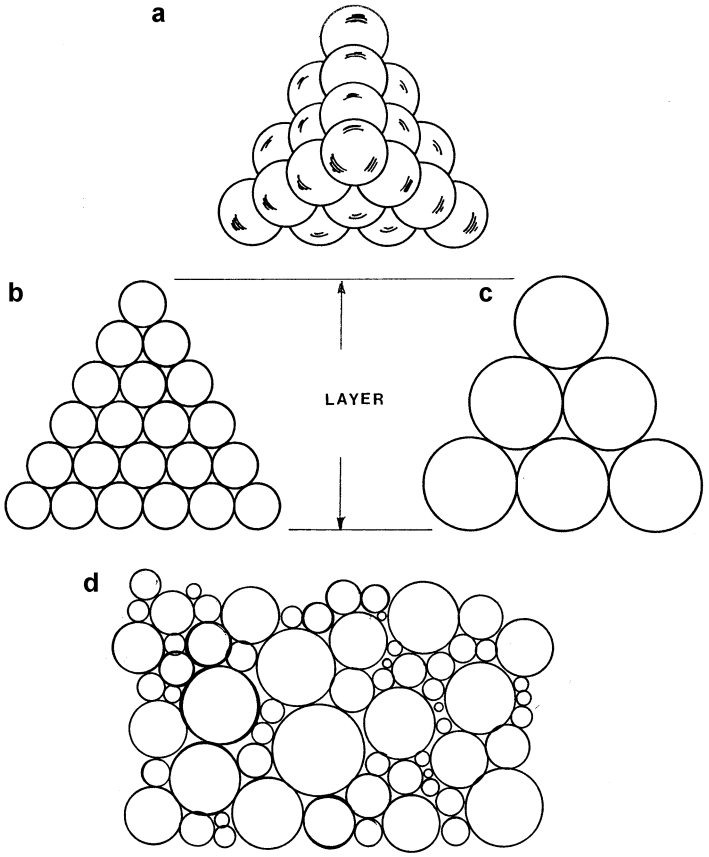


Fig. 3 Idealized powder packing arrays: unisized (**a**–**c**) and distributed particle sizes (**d**). (**a**) shows top view section of (**b**) and (**c**). (**d**) can represent both the side (layer thickness) and top views, respectively

1950s, and today there are nearly 2000 systems worldwide. Electron beams interact differently from laser beams and electron beams can produce much higher energy densities than laser beams. Electron beam melting (EBM) systems, like electron beam welding systems, resemble classical electron optical configurations such as scanning or transmission electron microscopes although the beam currents are much higher and the lens systems are simpler.

Selective Laser Melting (SLM) Principles

Figure 1 illustrates the main features of the SLM system, which, except for an inert atmosphere requirement and a higher power laser, is the same as the SLS system in ► Fig. 10 of chapter "Rapid Prototyping Technologies: Solid Freeform Fabrication."

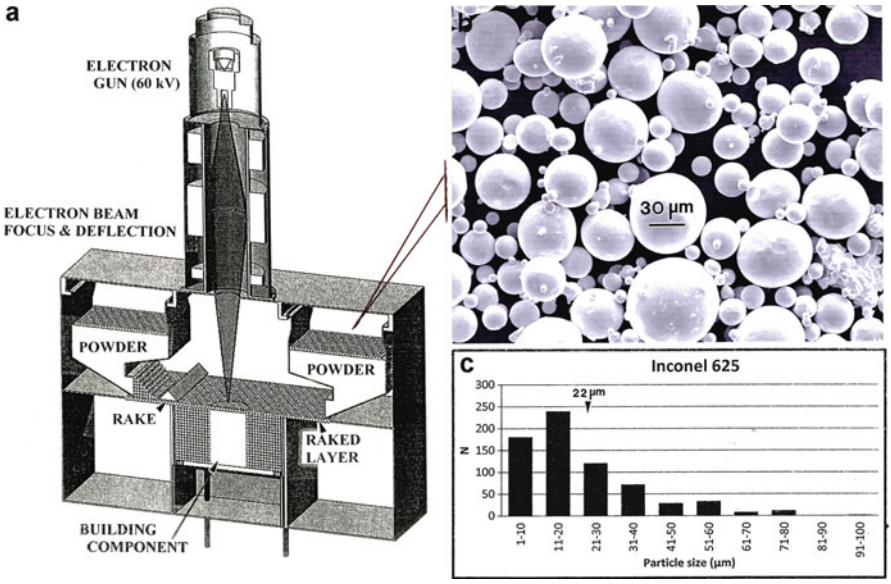


Fig. 4 EBM system schematic (**a**) with representative, pre-alloyed (Inconel 625), nickel-base superalloy powder example (**b**) and the corresponding particle size histogram representing particle size distribution corresponding to Fig. 3 (**d**). (**c**) The average particle size is 22 µm (*arrow*) in (**c**)

The inert gas employed in SLM fabrication of metal or alloy powders is usually purified argon (Ar) or nitrogen (N_2 or N2). The corrective lens shown at (3) in Fig. 1 ensures the beam is round as it scans the powder bed on the build platform (5) and maintains the beam velocity proportional to the angular velocity of the scanning mirror at (2).

As shown ideally in Fig. 2, the scanning beam selectively melts the rolled powder bed layer which correspondingly melts to the previous melted and cooled layer. As in the SLS process (► Fig. 10 of chapter "Rapid Prototyping Technologies: Solid Freeform Fabrication"), the powder bed (and build) is slightly heated. The laser beam, or beam energy, is absorbed in the powder layer in some correspondence with the particle size and packing or packing density. The adsorption for metals like copper powder in Fig. 2b is a factor 15 greater for CO_2 laser energy (at 10.6 µm wavelength) than YAG (at 1.06 µm wavelength) laser energy (0.1 vs. 0.015; Duley 1983). Particle packing also influences thermal conductivity, increasing for more closely packed powder as illustrated in Fig. 3. A porous powder bed also encourages reflectivity of the beam, decreasing process efficiency. Multi-reflection phenomena also cause radiation transport in deeper powder layers and can cause melt bonding which is often observable in the fabricated metal or alloy product.

In the simplest case, the energy input required to initiate melting along a scan vector with constant beam diameter, d, and scan line spacing depends on the laser power, P_L, and the scanning velocity, v_s:

Fig. 5 Acram A2 EBM system

$$E_i = P_L / v_s \tag{1}$$

Correspondingly, the minimum melt (or fusion) energy for the powder (particle) layer bed is

$$E_{min} = \left(C_p \Delta T + C_M\right) \rho v, \tag{2}$$

where C_p is the heat capacity at constant pressure, C_M is the latent melt energy, ρ is the powder particle density, V is the average particle volume ($4\,\pi\, r^3/_3$; where r is the particle radius), and ΔT is the temperature rise needed for melting to occur. From Eq. 2, the process efficiency can be expressed as

$$\eta = E_{min} / \rho d v_s, \tag{3}$$

where ρ is the density and d is the laser beam (focused) diameter or spot size.

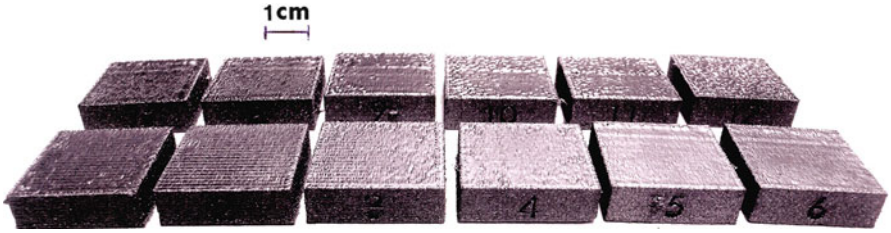


Fig. 6 Ti-6Al-4 V focus (offset) optimization test block sequence for EBM processing. Blocks are denoted (1–12). Note variations of surface roughness associated with melt scan lines

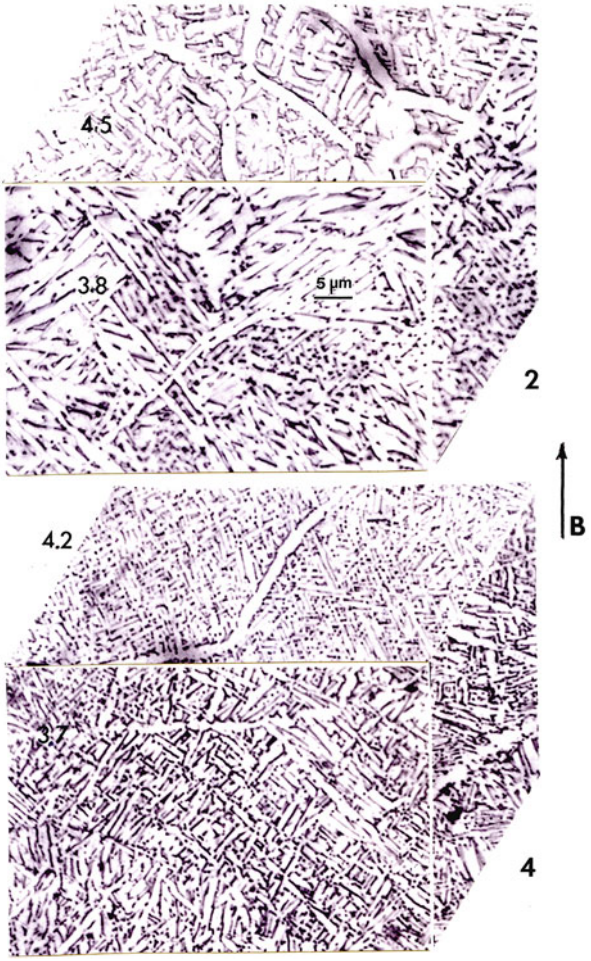


Fig. 7 Ti-6Al-4 V 3D light microscope compositions corresponding to focus offset test block numbers in Fig. 6. The build direction is noted by *arrow B*. Compositions correspond to test block sections in Fig. 6. Microindentation hardness numbers are shown in GPa. The softest microstructures correspond to test blocks (*6*) and (*12*) (Fig. 6)

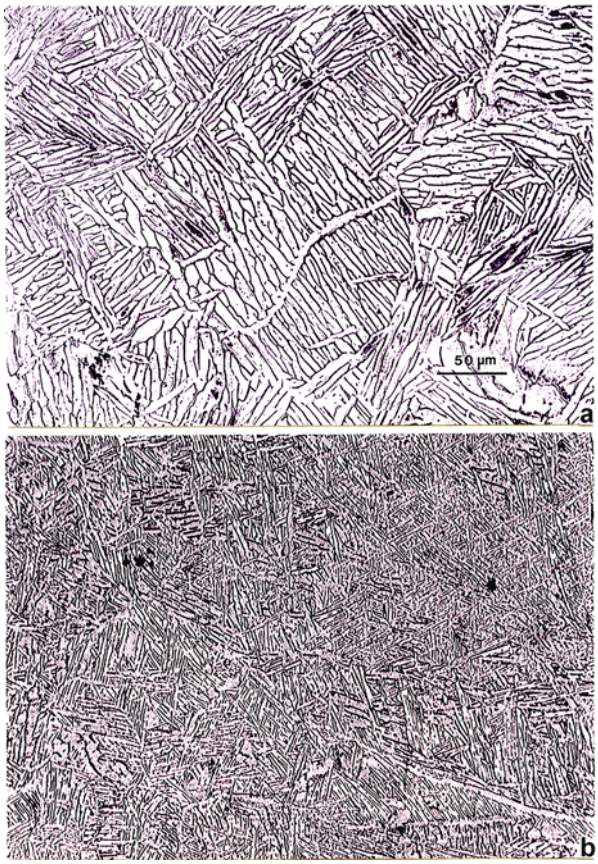


Fig. 8 Lenticular (d-hcp) phase variation in EBM-fabricated Ti-6Al-4 V alloy at different build conditions. (**a**) 6 mA beam current, 300 m/s melt scan speed. 3.4 GPa microindentation hardness. (**b**) 3 mA beam current, 400 m/s melt scan speed. 4.3 GPa microindentation hardness

Electron Beam Melting (EBM) Principles

Arcam AB, Sweden, adapted electron beam technology for SFF in the late 1990s. Although conceptually similar to SLM, EBM systems shown schematically in Fig. 4a generate a high energy electron beam in a standard electron gun configuration operating at 60 kV accelerating potential. Beam currents in the tens of milliamperes are used to create beam energies comparable to laser beam energies. However, the scan speeds for the EBM system are orders of magnitude greater than laser melting systems as shown in Fig. 1.

The electron beam interaction with the powder bed is complex and fundamentally different from laser beam interaction. This is due to the change and momentum (or kinetic energy) of the electrons. Interactions of the electron beam are also correlated with the powder properties (size, size distribution, density, chemistry,

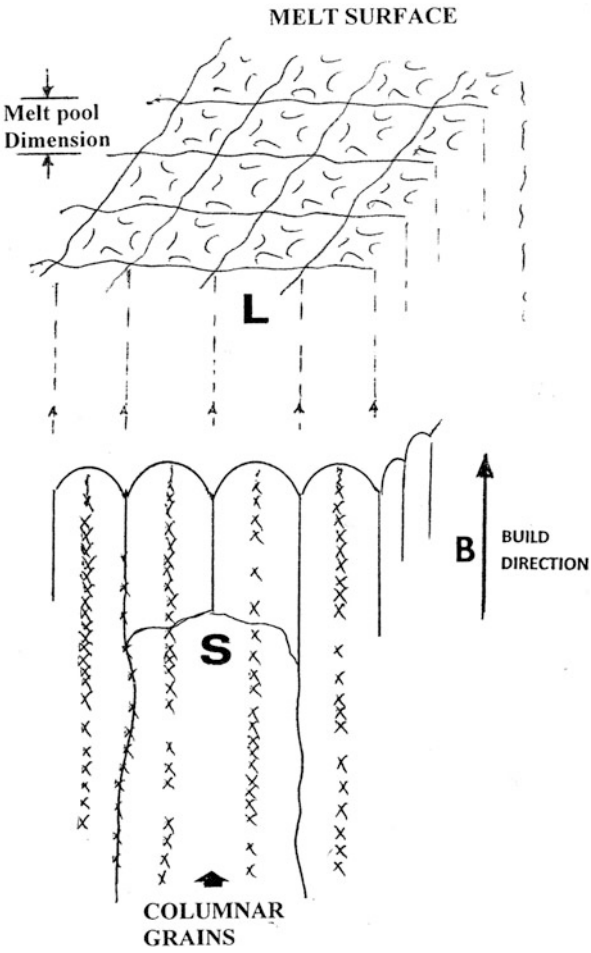


Fig. 9 SLM or EBM process/fabrication features: the formation of the melt layer composed of discrete melt pools (*L*). Columnar grains and melt pool center-oriented microstructures are shown in the solid (*S*) build

etc.) which are illustrated physically in Fig. 4b, c for a nickel-base superalloy: Inconel 625 (66 % Ni, 21 % Cr, 9 % Mo, 4 % Nb, 0.4 % Fe, traces of C and Ti; in weight percent). The associated electrostatic and electrodynamic forces are also influenced by the electrical resistance (or corresponding conductance) of the powder which is dependent upon the particle size and size distribution (Fig. 3) which controls packing or contact. This also influences thermal conductivity as noted above for SLM. These complex interactions often give rise to powder dispersal on contact with the electron beam. To avoid this problem, the powder bed in EBM (Fig. 4a) is prescanned to induce temperatures which just begin to sinter the particles, usually $\geq$0.4 T_M (where T_M is the melting temperature). This prescan or pore-heat scan is accomplished at high beam current and at very fast scan speeds (10^3mm/s), roughly 10^2 more rapid than the melt scan which is also accomplished

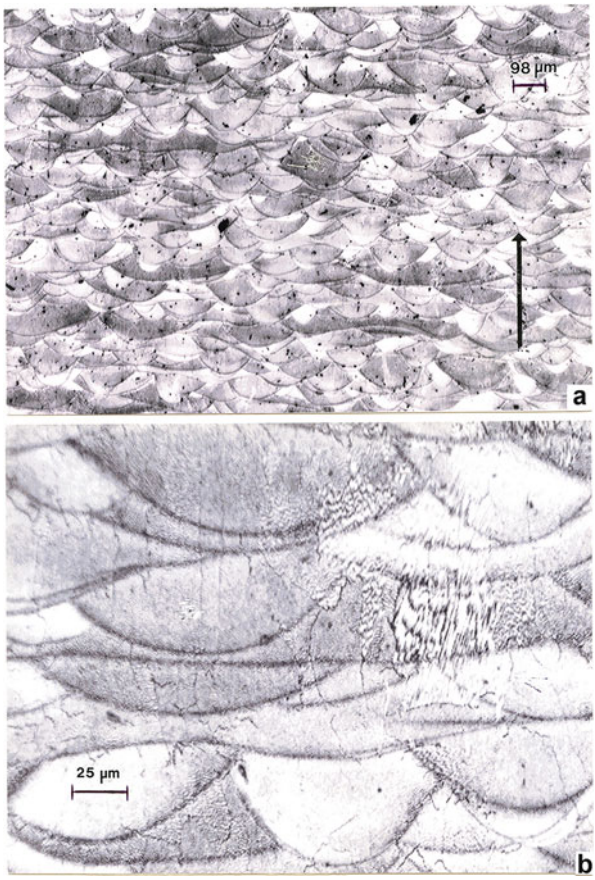


Fig. 10 Melt pool banding corresponding to alternating melt scan vectors in SLM fabrication of pre-alloyed Inconel 718 powder (53.5 Ni-19 Cr-18.3 Fe-5Nb-3Mo-1 Ti-0.43 Al). (**a**) Low magnification (**b**) higher magnification. The build direction is shown by *arrow* in (**a**)

at reduced beam current. Consequently, these adjustments in beam power and scan speed control the energy input as implicit in Eq. 1. The preheat scanning can often involve up to a dozen cycles where the beam is scanned in alternating x-y directions or scan vectors across the powder bed (or raked layer) as implicit in Fig. 4a.

Figure 5 shows the Acram A2 EBM system, which in 2013 was the only commercially available EBM system. The left section of the unit in Fig. 5 is the control system which utilizes an STL – CAD file to rake the powder bed layer, preheat the powder, and selectively melt each preheated layer according to specified build strategies invested in the CAD program. This includes optimized build parameters such as beam current, beam spot size, beam vector (or path) spacing, scan rate and sequencing (x-y) (or x, y in alternating layers, etc.), melt scan rate and beam current, etc. Much of this build strategy development and optimization (parameter optimization) is done by trial-and-error experimental runs or abbreviated builds and build geometries. Beam focus, depending to some extent on powder size or size distribution as implicit in Eqs. 2 and 33, can be selected from a series of

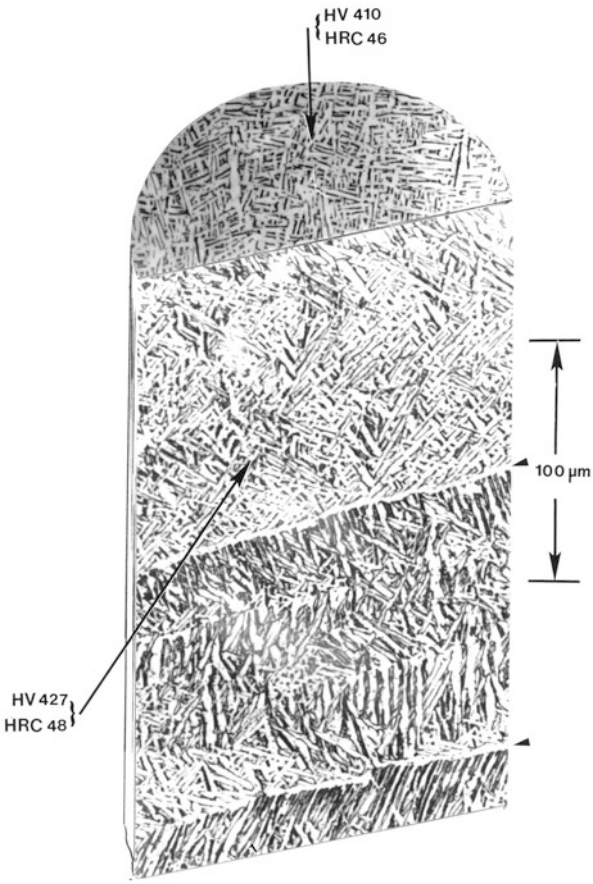


Fig. 11 EBM 3D-optical (light) microscope composition representing a section of a Ti-6Al-4 V fabricated product, showing layer striations or bands (*arrows* at *right*) representing layer thicknesses of ~100 µm. Vickers microindentation (*HV*) and Rockwell C-scale hardnesses (*HRC*) are shown

so-called focus offset builds where the surface melt and melt scan appearance selected also often based on examination of the residual and associated microstructure. Figure 6 illustrates a dozen such focus offset build trials, while Fig. 7 shows some corresponding microstructures in 3D optical metallograph constructions for a Ti-6Al-4 V alloy, with a melting point of 1,570 °C.

In this optimization sequence (Figs. 6 and 7), focus offset number (6) was selected. These beam focus conditions also involve variations in beam current as well as spot or beam size (d) variations, where in effect the beam focus is occurring at some distance from the powder bed surface and within the specified build layer dimension, which, as noted previously (Fig. 3), will depend on the powder particle size and size distribution. And while Fig. 7 shows only small variations in the α-lenticular grain as phase sizes in the Ti-6Al-4 V alloy with focus offset numbers, it is impossible to control or vary the grain size or the microstructure through additional variations or additions to the build strategy, including multiple melt scans or post-melt heating which can induce grain or phase growth or phase

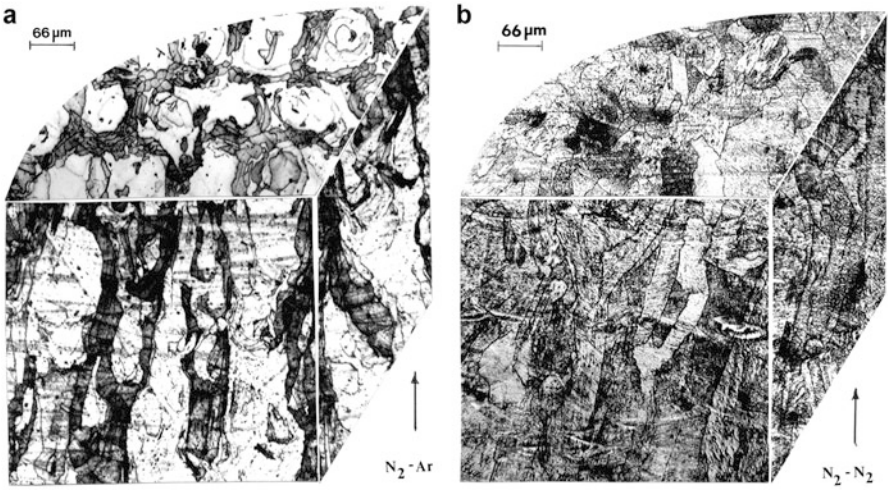


Fig. 12 Comparison of 3D – light microscope compositions for SLM-fabricated 17–4 PH components. (**a**) Pre-alloyed, 17–4 PH powder atomized in N_2 and fabricated by SLM in an argon (*Ar*) environment (N_2/Ar). (**b**) Pre-alloyed 17–4 PH powder atomized in N_2 and fabricated by SLM in a nitrogen (N_2) environment (N_2/N_2)

changes. Figure 8 shows grain size (or α-phase size) changes with beam current (beam power or energy density: energy/unit area of beam spot). We will examine these issues in more detail later, especially in the context of build (product) cooling rate as it is affected by the product dimensions or mass. Large or tall parts with large mass remain hot especially the top because of heat conduction. Similar phenomena occur for SLM-fabricated components as well and especially since the beam scan rates are lower and the inert gas environments provide more efficient convective cooling than the vacuum of the EBM system.

SLM and EBM Application Examples: Materials Issues

The evolution of microstructure in SLM or EBM fabrication is more complex than growth-related microstructure evolution for castings or directional crystal growth or even more conventional solidification processes. This is because the melt alternates in the layer building process and occurs in isolated or semi-isolated melt pools especially for x-y beam scanning, which produces a Gaussian-like energy and heating profile within the melt pool volume. The fact that isolated melt pools are formed on previously melted and cooled underlayers which may approximate the same melt pool dimension might prompt some association with portioned epitaxy. This concept might be visualized from the schematic view provided in Fig. 9. The melt pool dimension shown in Fig. 9 depends primarily on the scan line width which also is related to the spot size or beam diameter at the powder layer surface.

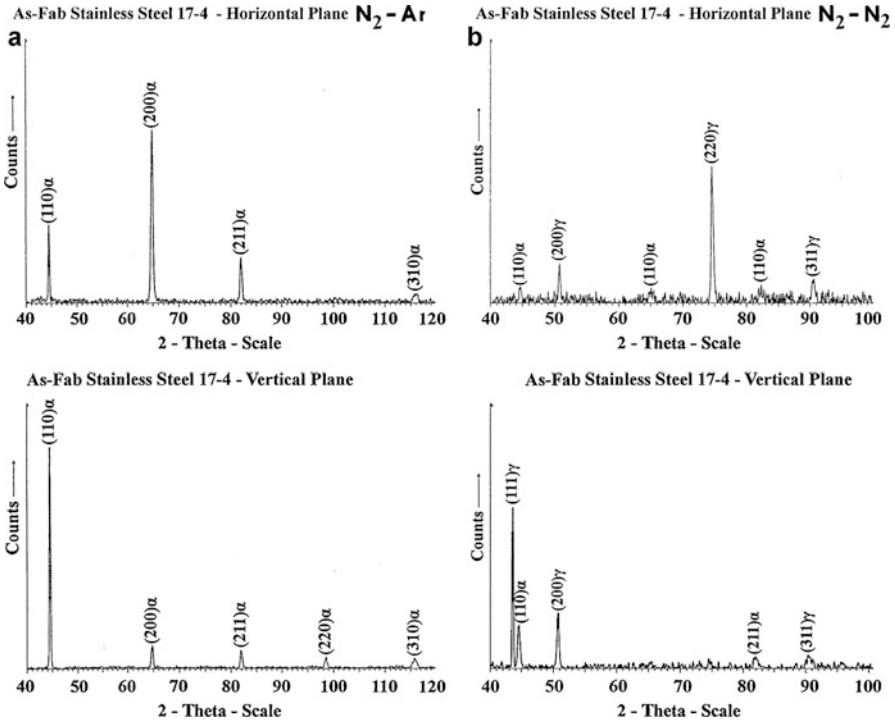


Fig. 13 Comparative XRD spectra corresponding to the SLM-fabricated 17–4 PH components in Fig. 12. (**a**) Fabricated in argon (N_2–Ar). Horizontal and vertical reference planes correspond to 3D composition in Fig. 12a. (**b**) Fabricated in nitrogen (N_2–N_2). Horizontal and vertical reference planes correspond to 3D composition in Fig. 12

Depending on the beam energy or energy density, the heat flux or heat profile will change, providing a temperature variation between the center of the melt pool and its edges where it meets other melt pool profiles as illustrated in Fig. 9. Nucleation and growth will, under some thermodynamic or thermokinetic circumstances imposed by the build strategy, promote directional growth and associated microstructures which may be altered or interrupted at the solid/melt interface, corresponding to the layer thickness. These features are more pronounced for SLM in contrast to EBM as a consequence of the variations and complexities of energy deposition from the beam (laser vs. electron) and the beam energy (or power) and scan rate (Eq. 1). This is illustrated comparatively in Figs. 10 and 11 for the SLM building of an Inconel 718 component and the EBM building of a Ti-6Al-4 V component, respectively. The layer or layer transition region in each case represents a specific microstructural and microchemical regime, which can be the same or different from the layer itself. It can also be suppressed or eliminated by selecting appropriate build parameters in either SLM or EBM.

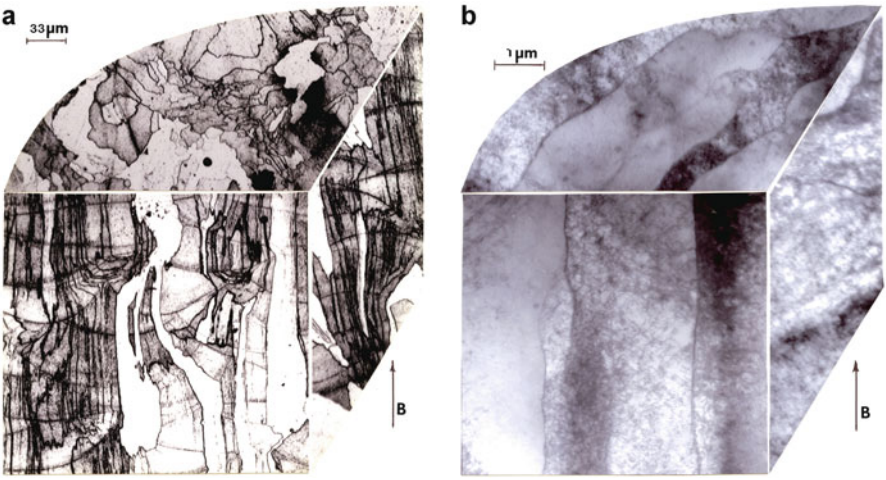


Fig. 14 Comparison of 3D – light microscope (**a**) and transmission electron microscope (**b**) compositions for 17–4 PH pre-alloyed powder atomized in argon and fabricated by SLM in an argon gas environment

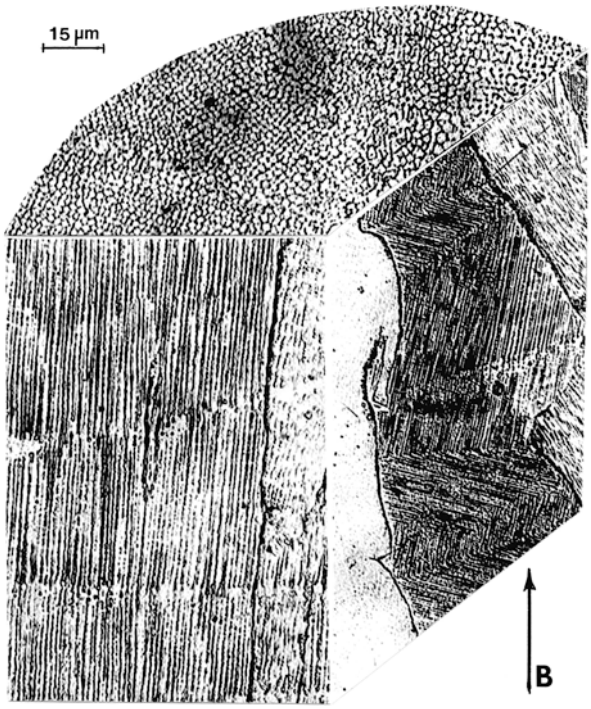


Fig. 15 3D light microscope image composition for EBM-fabricated copper. Irregular columns of Cu_2O precipitates are formed. *B* denotes the build direction

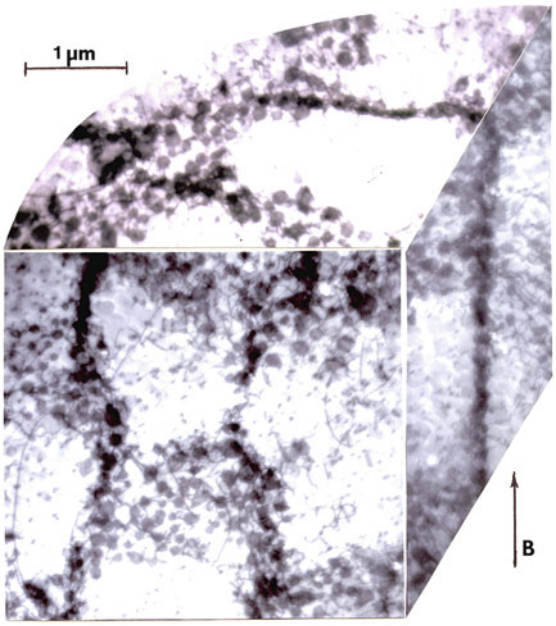


Fig. 16 3D-TEM image composition representing a magnified section in Fig. 15 showing Cu_2O precipitates intermixed with dislocations and dislocation tangles. Small dots in the clear regions or cell-like centers are small dislocation loops. *B* denotes the build direction as in Fig. 15

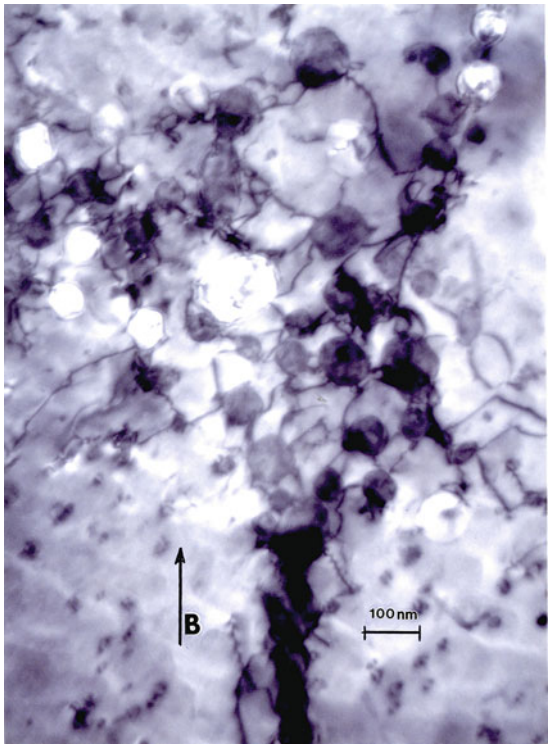


Fig. 17 Magnified TEM image for Cu_2O precipitates and dislocation microstructure corresponding to Fig. 16 vertical reference plane, parallel to the build direction indicated by *B*

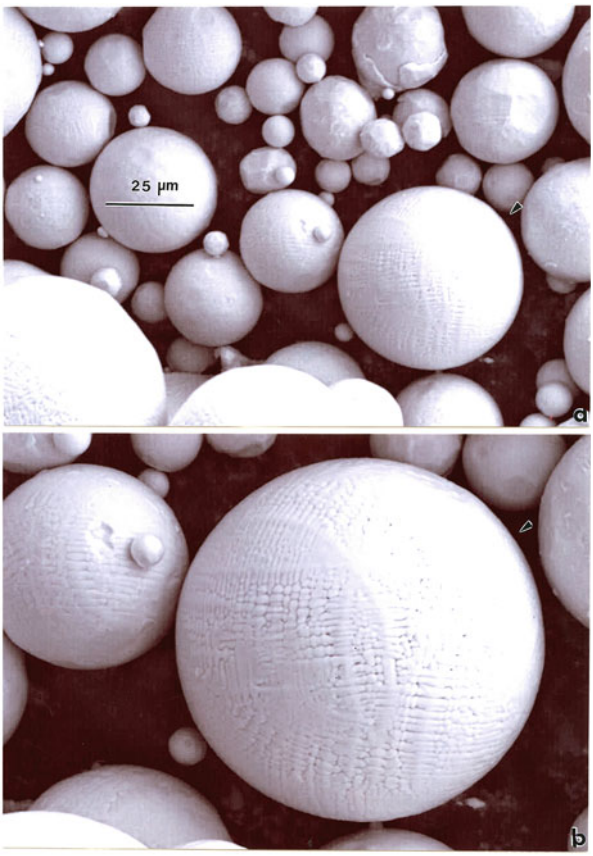


Fig. 18 SEM view of atomized alloy 625 (Ni-base superalloy) powder (**a**). (**b**) shows magnified view of particle marked with small arrow in (**a**). Note microdendrite structure

There are aspects of this observable on comparing Fig. 12a, b with Fig. 10. The melt pool layers are less apparent in Fig. 12b than in Fig. 10. In addition, the 17–4 PH stainless steel builds represented in Fig. 12a, b represent another unique feature of SLM fabrication as it relates to the gas environment (Fig. 1). Figure 12a shows irregular, columnar-like α-martensite (bcc) grains oriented generally in the build direction (heavy arrow). The 17–4 PH pre-alloyed powder (17.5 Cr, 4 Ni, 4 Cu, 1 Mn, 0.5 Nb, bal. Fe in wt. %) used in the building of components represented in Fig. 12a was gas atomized, or rapidly solidified, in purified nitrogen (N_2), while the SLM building was performed in an argon (Ar) environment (designated N_2–Ar) using an EOS (Germany) M 270 SLM machine with a 0.2 kW laser source. By contrast, using this atomized powder by SLM building in a nitrogen (N2) environment produced a correspondingly different ν-fcc (austenitic) microstructure as illustrated for comparison in Fig. 12b. The → phase transition implicit in the microstructure comparisons in Fig. 12 is confirmed in the comparative and corresponding XRD spectra shown in Fig. 13a, b. Murr et al. (2012) have described this α-Fe → γ-Fe transition by SLM building in an Ar environment in

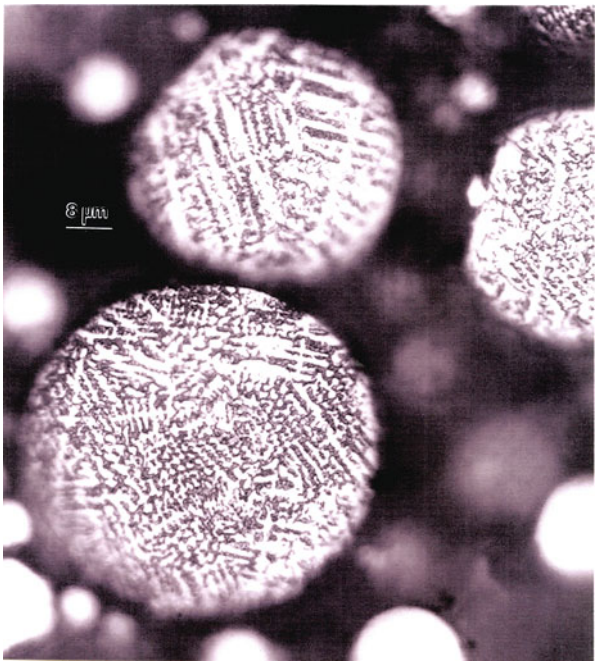


Fig. 19 Light microscope cross-section view of microdendrite structure in atomized alloy 625 powder particles in Fig. 18b

Fig. 20 3D-light microscope image composite showing columns of γ″ (Ni_3Nb) nanoprecipitates in EBM-fabricated alloy 625 (Inconel 625: Ni-base superalloy). Grain boundary is noted at GB. The arrow to right of GB denotes the build direction

Fig. 21 Secondary electron-SEM image for the γ″ precipitate columns in Fig. 20. The γ″ precipitate plates are observed to be coincident with the Ni–Cr (fcc) matrix {111} plane designated by *arrows a* and *b*. *B* denotes the EBM build direction

contrast to an N_2 environment as a result of the constitutional cooling in N_2 in contrast to Ar because the thermal conductivity of N_2 gas is ~40 % greater than that of Ar gas over a wide temperature range. As observed in Fig. 13a, the build is completely α-Fe (martensite) and is highly magnetic, while only remnants of α-Fe are retained in the N_2-fabricated 17–4 PH product in Fig. 13b, which is nonmagnetic. In addition, the residual microindentation hardness associated with the α-Fe magnetic product in Fig. 12a was ~3.9 GPa while that for the γ-Fe product in Fig. 12b was ~2.2 GPa (Murr et al. 2012). Figure 14a shows SLM-fabricated 17–4 PH component using pre-alloyed powder atomized in argon and fabricated in argon gas as in Fig. 12a. The XRD spectrum was identical to Fig. 13a, indicative of α-Fe. Figure 14b shows the dislocation microstructure composing α-Fe laths oriented in the build direction (B in Fig. 14b). It can also be noted in Fig. 13a, b that there is recognizable texturing in both the horizontal reference plane perpendicular to the build direction and the vertical reference plane parallel to the build direction (Fig. 12a): [200] α versus [110] α. In contrast, the γ-Fe (austenitic) material represented in Fig. 13b shows correspondingly [220] γ versus [111] γ. This phenomenon can be compared with the schematic perspective for build microstructures or microstructure evolution illustrated in Fig. 9.

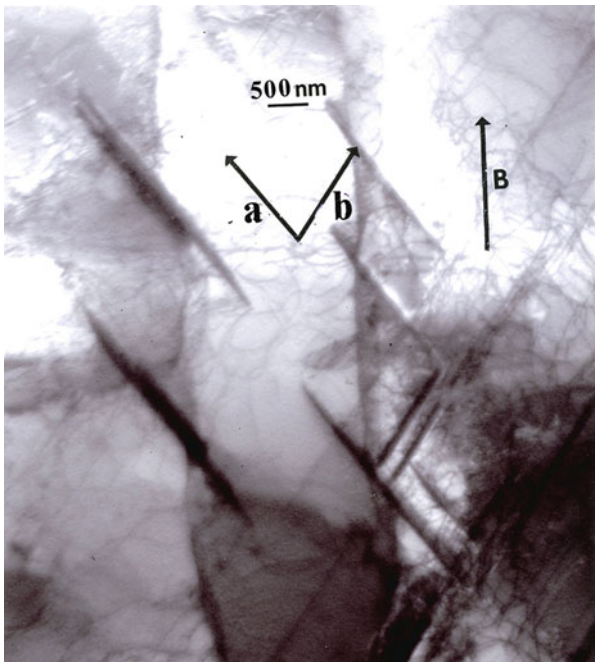


Fig. 22 TEM bright-field image representing a magnified section in Fig. 21 showing the γ″ precipitates coincident with Ni–Cr fcc matrix {111} planes indicated by trace directions marked *a* and *b*. These correspond to *b* and *a* in Fig. 21. Dislocations are also evident surrounding the γ″ precipitates

Although many applications of EBM fabrication in the subsection entitled "► Implant Materials and Structures" under section "Biomaterials" have already been illustrated, a few novel examples will be added to compliment the SLM examples provided in Figs. 12, 13, and 14. These examples begin with a somewhat novel example of EBM – process generated, columnar precipitate microstructure in copper illustrated in Fig. 15. The copper, as shown in Fig. 15, contained Cu_2O precipitates as a consequence of reaction with traces of oxygen during atomization in argon. During EBM fabrication, the Cu_2O precipitates became entrained in the melt pool along with additional precipitates as illustrated schematically in Fig. 9 (Ramirez, et al. 2011). Figures 16 and 17 show the details of the Cu_2O precipitates, forming complex and interactive microstructures with dislocations and dislocation loops. In Figs. 15, 16, and 17, the build direction is noted by B, and it can be observed in Fig. 15 that the columnar, precipitate (Cu_2O) architecture is interrupted by the scanning sequence where it can be noted at A and, within the columnar grain in the right face of the 3D composition, that the Cu_2O precipitate columns make angles of 90° with one another. This is a consequence of the x-y scan variation within the build. It can also be noted in Fig. 17 that some of the precipitates etch out in the etchant used to create the TEM thin section.

► Figures 20 and ► 21 of chapter "Implant Materials and Structures" have previously illustrated similar columnar precipitate ($Cr_{23}C_6$) architectures in the EBM fabrication of Co–Cr superalloy: compare ► Fig. 20 of chapter "Implant

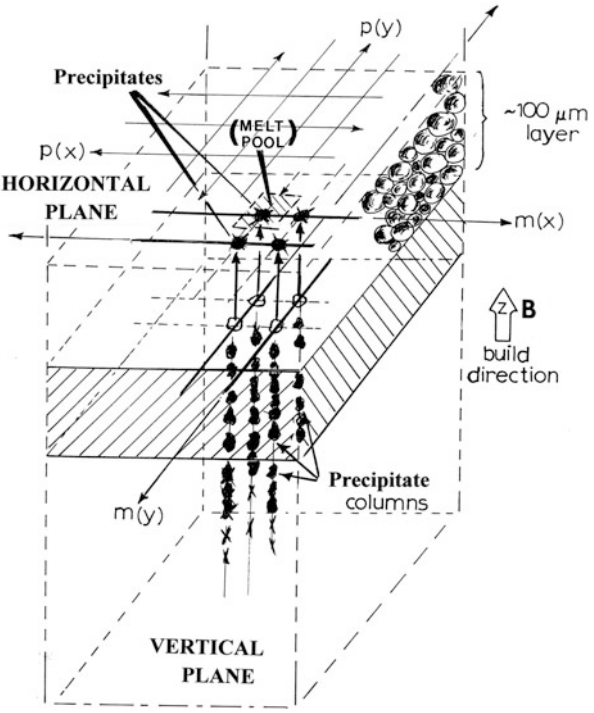


Fig. 23 Schematic representation for columnar precipitate development in EBM fabrication. The horizontal and vertical planes denote the reference planes characterizing the 3D image compositions in Figs. 15 and 20 (Adapted from Gaytan et al. 2010)

Materials and Structures" with Fig. 15 in particular. In the sequence of SEM–light microscope–TEM images in Figs. 18, 19, 20, 21, 22, and 23, we illustrate an additional example of this phenomenon which characterizes EBM processing of many alloys exhibiting precipitation thermodynamics. Figs. 18 and 19 show examples of the structures and microstructures for Inconel 625, a nickel-base superalloy: nominally 61Ni–22Cr–9Mo–3.5Nb–3.2Fe–0.02C–0.7Mn, Si, Al, Ti, in weight percent. As illustrated in Fig. 18b, this rapidly solidified powder is composed of microdendritic structure which is shown in the inside view of the particles in the light microscope images of Fig. 19. Figure 20 illustrates the typical EBM build microstructure and microstructural architecture composed of columnar grains (GB) and columnar γ" (Ni_3Nb) bct (a = 3.6 Å, c = 7.4 Å) precipitates. These are spaced within the melt pool columns illustrated schematically in Fig. 9. A comparison of these γ″ precipitate columns shown in Figs. 20 and ▶ Fig. 22 of chapter "Laser and Electron Beam Melting Technologies," and in the detailed TEM image in Fig. 22, illustrates that these γ" precipitates are platelets coincident with the Ni–Cr (fcc) matrix {111} planes (Murr et al. 2011).

Looking again in retrospect at ▶ Figs. 20 and ▶ 21 of chapter "Implant Materials and Structures" as well as Figs. 15 and 16, Figs. 16, 17, and 18, and Fig. 9, one might generalize this microstructure/columnar architecture

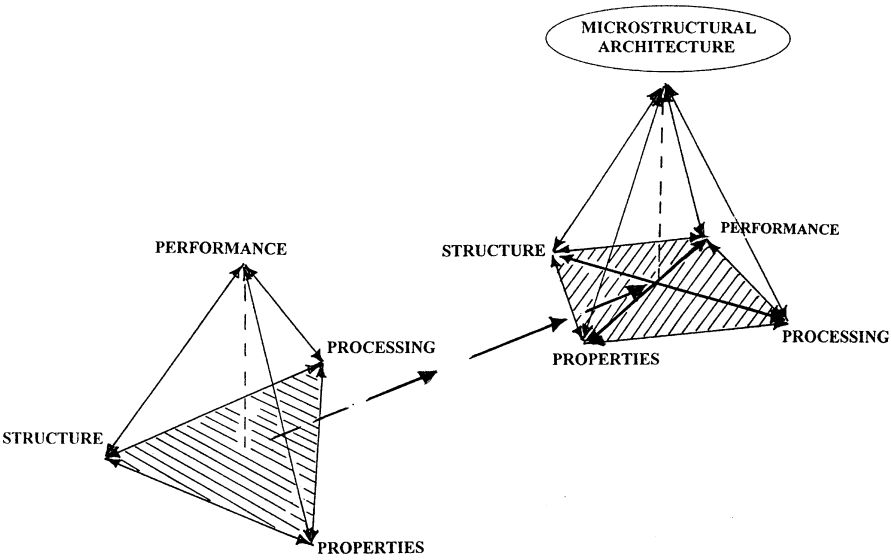


Fig. 24 New materials science and engineering paradigm shift: from the structure-properties-processing-performance tetrahedron to a pyramid with microstructural architecture at the top vertex

development in EBM processing as illustrated in Fig. 19. These features have also been observed in the SLM fabrication of Inconel 718 nickel-base superalloy as well, where columns of γ' precipitates form coincident with the Ni–Cr (fcc) matrix {100} planes (Amato et al. 2012). In all of these examples, the columnar precipitates, having different crystal structures and compositions, are nanosize. Consequently laser and electron beam fabrication can somewhat uniquely control constituent (nanoconstituent) architectures. While aspects of this phenomenon have been employed in a variety of processing regimes for decades, such as precision, directional casting of turbine blades as illustrated in ▶ Fig. 1c, d of chapter "Examples of Directional Crystal Structures: Gas Turbine Component Applications in Superalloys," the precision of nanoscale engineering which is beginning to emerge would seem to imply an expansion of the traditional materials science and engineering paradigm involving structure, properties, processing, and performance to include the manipulation of microstructural architecture. This is illustrated in the diagrams of Fig. 24. Conceptually, this illustrates, along with the examples cited above, that materials science and engineering, in a broad sense, is entering an era in which materials properties are not only obtained or manipulated by processing routes but also by controlling or manipulating the architecture of its constituents, especially those at the nanoscale. These nanoscale issues will be explored in more detail in chapter "▶ 3D and Multidimensional Materials Science."

References

Amato K, Murr LE, Gaytan SM, Martinez E, Shindo PW, Hernandez J, Collins S, Medina F (2012) Microstructures and mechanical behavior for Inconel 718 fabricated by selective laser melting. Acta Mater 60:2229–2239

Duley WW (1983) Laser processing and analysis of materials. Plenum, New York

Gaytan SM, Murr LE, Martinez E, Martinez JL, Machado BI, Ramirez DA, Medina F, Collins S, Wicker RB (2010) Comparison of microstructures and mechanical properties for solid and mesh cobalt-base alloy prototypes fabricated by electron beam melting. Metall Mater Trans A 41A:3216–3227

Murr LE, Martinez E, Gaytan SM, Ramirez DA, Machado BI, Shindo PW, Martinez JL, Medina F, Wooten J, Ciscel D, Ackelid U, Wicker RB (2011) Microstructural architecture, microstructures, and mechanical properties for a nickel-base superalloy fabricated by electron beam melting. Metall Mater Trans A 42A:3491–3508

Murr LE, Martinez E, Hernandez J, Collins S, Amato KN, Gaytan SM, Shindo PW (2012) Microstructures and properties of 17-4 PH stainless steel fabricated by selective laser melting. J Mater Res Technol 1(3):167–177

Ramirez DA, Murr LE, Martinez E, Hernandez DH, Hernandez JL, Machado BI, Medina F, Wicker RB, Frigola P (2011) Novel precipitate-microstructural architecture developed in the fabrication of solid copper components by additive manufacturing using electron beam melting. Acta Mater 59:4088–4099

3D and Multidimensional Materials Science

Contents

Abstract

Many aspects of materials science and engineering structures are conveniently and effectively rendered by 3D imaging and related or associated properties. As these structures evolve systematically with time, such systematic renderings are characterized as 4D. Examples of these concepts are presented in this chapter.

Introduction

The reader should appreciate the examples of 3D materials science provided in the many 3D image compositions shown for light microscope and TEM image perspectives: ▶ Fig. 20 of chapter "Implant Materials and Structures" and ▶ Figs. 11–13, ▶ 15–17, ▶ 20–23, etc., in Chapter "Laser and Electron Beam Melting Technologies". Of course these perspectives can be expanded to sequences of microstructure or microstructural architecture variations with time or temperature, or time and temperature, or even other measured physical quantities such as strength, hardness, conductivity, etc.

Narrative

Even systematic changes involving microstructural or architectural evolution with time, temperature, or other processing variables can be systematically observed by sequences of multidimensional observations like frames in a

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_41

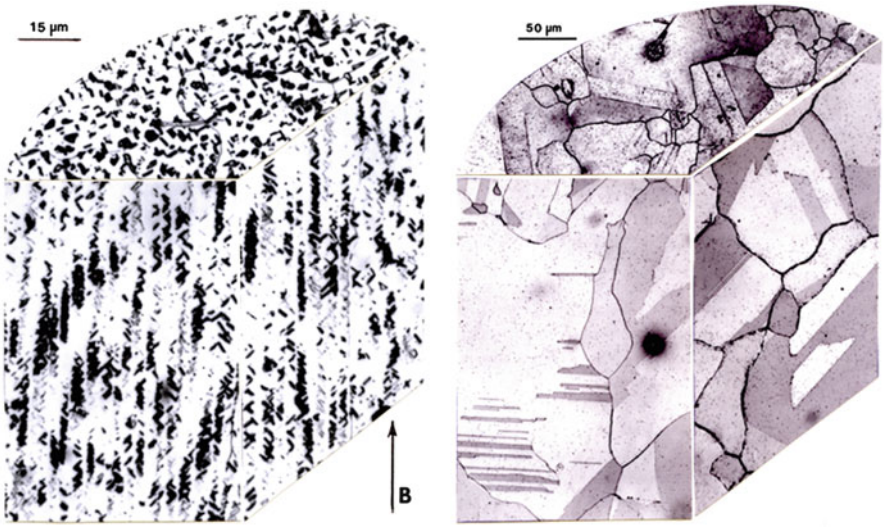


Fig. 1 3D light microscope compositions showing microstructure evolution end points with temperature and time (T(t)) representing multidimensional materials science. γ″ precipitate columns for EBM-fabricated alloy 625 evolve into a mostly equiaxed fcc Ni-Cr grain structure with {111} coincident annealing twins and $NiCr_2$ globular nanoprecipitates after heating 4 h at 1,120 ° C

digital or analogue image stream. In dimensional analysis, physical quantities which are readily measurable involve basic dimensions – mass, length, time, temperature, electrical current, etc. – although classical multidimensional concepts involve Cartesian coordinates (x, y, z) representing 3D and time, which is often expressed as 4D. Figure 1 illustrates a simple example of microstructural evolution and microstructural architecture variation with time and temperature (both constant) for the EBM-fabricated Inconel 625 superalloy described in ▶ Figs. 20–22 in chapter "Laser and Electron Beam Melting Technologies". In ▶ Fig. 2 of chapter "Laser and Electron Beam Melting Technologies," the γ″ (bct) precipitate (nanoprecipitate) architecture dissolves after 4 h at 1,120 °C (0.1 GPa pressure) HIP (hot isostatic pressing) in purified argon. This leads also to the formation of equiaxed grain structure containing annealing twins as a consequence of the low stacking-fault free energy Ni-Cr fcc matrix. Careful observation of the HIP-annealed microstructure evolution in Fig. 1 reveals tiny (nano), spherical or globular precipitates of $NiCr_2$ (laves) phase (Murr et al. 2011). Of course heat treatment at this temperature at different times (starting from t = o) would reveal the details of the microstructure evolution, including grain growth, etc. Similar snapshots of microstructure evolution are shown in comparing Figs. 2 and 3 which show the $Cr_{23}C_6$ precipitate columns in the Co-base superalloy fabricated by EBM in ▶ Fig. 21 of chapter "Implant

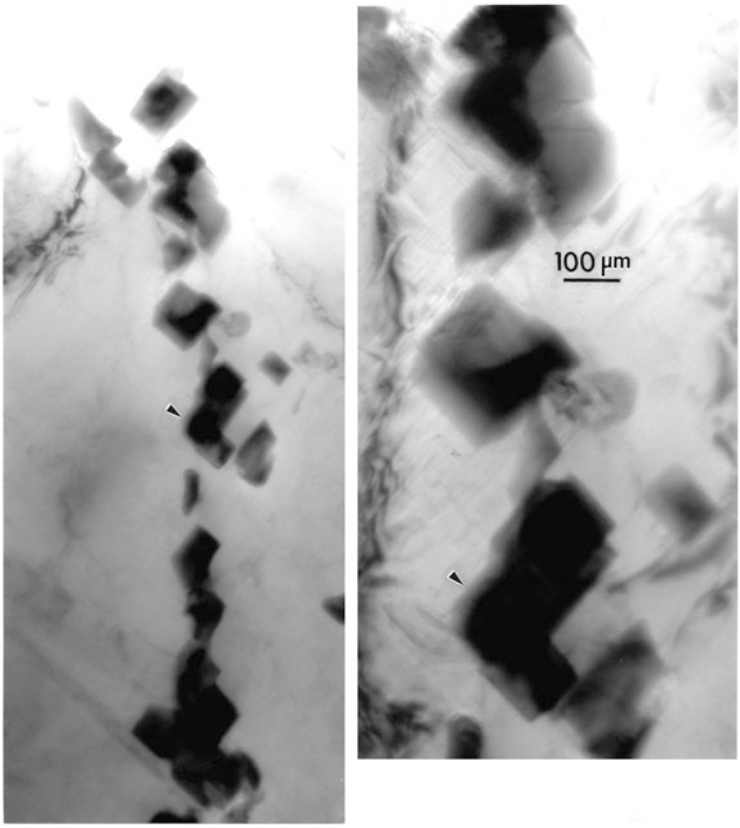


Fig. 2 TEM bright-field images showing $Cr_{23}C_6$ carbide precipitate columns corresponding to EBM-fabricated Co-base superalloy in ▶ Figs. 20 and ▶ 21 of chapter "Implant Materials and Structures". The right image is a magnified view of the left image referenced by the *arrow*. The cubic $Cr_{23}C_6$ precipitates have lattice parameter at – 10.6 Å. Note precipitate size is ~100 Å

Materials and Structures," which dissolve on annealing (or HIPing) at 1,200 °C for 4 h (at a pressure of 0.1 GPa in argon) and give rise to dense arrays of intrinsic stacking faults coincident with the Co-Cr fcc low stacking-fault free energy matrix as shown in Fig. 3 (Gaytan et al. 2010).

Of course these variations of microstructure with processing variables can be observed in the many examples of microstructures and microstructural evolution provided throughout this book. For example, compare ▶ Fig. 12 of chapter "Implant Materials and Structures" with ▶ Fig. 7 of chapter "Laser and Electron Beam Melting Technologies," or ▶ Fig. 7 of chapter "Laser and Electron Beam Melting Technologies" with ▶ Fig. 16 of chapter "Implant Materials and Structures," which are related to EBM processing of Ti-6Al-4V.

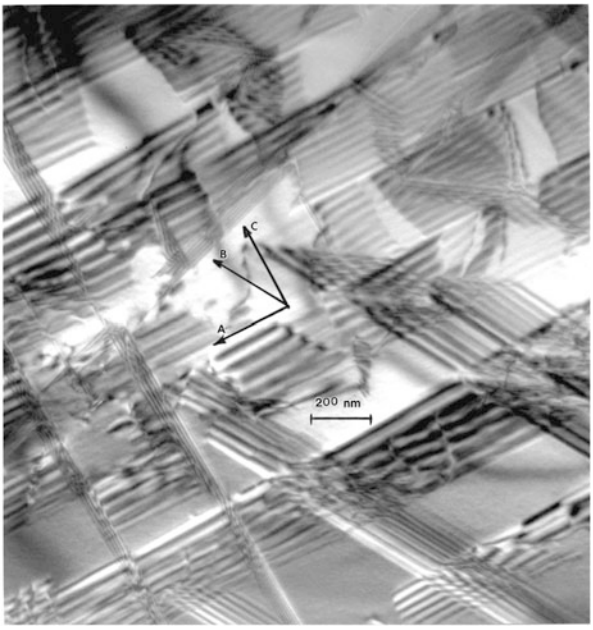


Fig. 3 TEM bright-field image showing intrinsic stacking faults in the EBM-fabricated product in 42.2 (and ▶ Figs. 20 and ▶ 21 of chapter "Implant Materials and Structures") after 4 h anneal at 1,200 °C. The grain surface orientation is (123). Traces of corresponding {111} planes are denoted *A*, *B*, *C*

References

Gaytan SM, Murr LE, Martinez E, Martinez JL, Machado BI, Ramirez DA, Medina F, Collins S, Wicker RB (2010) Comparison of microstructures and mechanical properties for solid and mesh cobalt-base alloy prototypes fabricated by electron beam melting. Metall Mater Trans A 41A:3216–3227

Murr LE, Martinez E, Gaytan SM, Ramirez DA, Machado BI, Shindo PW, Martinez JL, Medina F, Wooten J, Ciscel D, Ackelid U, Wicker RB (2011) Microstructural architecture, microstructures, and mechanical properties for a nickel-base superalloy fabricated by electron beam melting. Metall Mater Trans A 42A:3491–3508

Additive Manufacturing: Changing the Rules of Manufacturing

Contents

Abstract

Aspects of 3D printing and additive or layer manufacturing can be treated as modular manufacturing or modular components of manufacturing in the contemporary sense. Such modular manufacturing involves specialized product design and fabrication or product customization. These processes incorporate new materials along with new design strategies to achieve new performance features.

Introduction

For most of the twentieth century, manufacturing thrived on mass production involving interchangeable parts and product replication. It was rarely economical to manufacture a one-of-a-kind part or product, and most product manufacturing involved complex tooling and multiple, diverse operations or steps in creating a product. Multiple materials were also involved. Aside from casting technologies which often involve limited material removal, most manufacturing processes often fashion small components from larger precursors by cutting and machining: subtractive manufacturing. Such processes produce large waste fractions relative to the finished part: up to 90 % or more. Many wastes are often not conveniently or practically recyclable, and unless other, novel applications of waste products are found (often referred to as by-product synergism), cumbersome disposal strategies must be implemented.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_42

Narrative

Figure 1 provides a simple example of additive manufacturing advantages. The STL file model for an aircraft bracket is shown in Fig. 1a, while Fig. 1b illustrates the two bracket forms fabricated by EBM. Normally fabricated by CNC machining from a precursor block, roughly 90 % is waste, while the EBM-fabricated brackets require little finish machining or even simple tumble polishing to remove burrs. It can be observed in Fig. 1b that there is notable surface roughness in EBM fabrication which can be partly mitigated by adjusting the build strategy, but this may be of no consequence in some applications. For example, and as illustrated in Part X, custom-made (or patient-specific) implants of the same Ti-6Al-4V material as in Fig. 1b can benefit from surface roughness by encouraging bone cell ingrowth. On the other hand, SLM usually provides better surface finish and, for critical product fabrication, can be the preferred additive manufacturing strategy. Implicit in Fig. 1 is the fact that modifications to the CAD model to manufacture part variances are easily incorporated in the STL files, and no new tooling or tooling optimization is required.

It must be cautioned, however, that while powder-based additive manufacturing poses advantages in significant waste reduction, the cost of powder precursors is often significantly greater than equivalent precursor weight of conventional bar stock or billet forms, and powder availability becomes an important consideration. The limited build volume in SLM and EBM systems also poses limitations in specialty part fabrication as well as mass customization of even small parts, although as noted previously, both dental and other bone implant components have been, and continue to be, manufactured by EBM. As noted above, patient-specific designs can be inserted as CAD modifications requiring no tooling or tooling optimization.

Another advantage of additive manufacturing technologies, as illustrated throughout this chapter, and previous chapters, has been the ability to fabricate intricate, complex monoliths; even arrays of moving parts as illustrated in the FDM-manufactured monolith in Figs. 2 and 3.

The manufacture of intricate, internal channel structures for air and fluid flow in heat exchange systems of various application forms another viable application of additive manufacturing technologies, particularly SLM or EBM fabrication. One prominent example involves turbine blade and component manufacture, which is constantly increasing in volume and manufacturing efficiency worldwide. Aspects of turbine blade fabrication in the context of materials design and development in Chapter "▶ Examples of Directional Crystal Structures: Gas-Turbine Component Applications in Superalloys" have been addressed. Turbine component, and especially high-temperature blade operation (▶ Fig. 2 of chapter "Examples of Directional Crystal Structures: Gas-Turbine Component Applications in Superalloys") is somewhat limited by high-temperature metals and alloys, specifically their melting points and their high-temperature oxidation (or corrosion) which limits their operation and the ability to push operational efficiencies at continually elevated temperatures. One strategy in extending the operation of nickel-base and Co-base superalloys

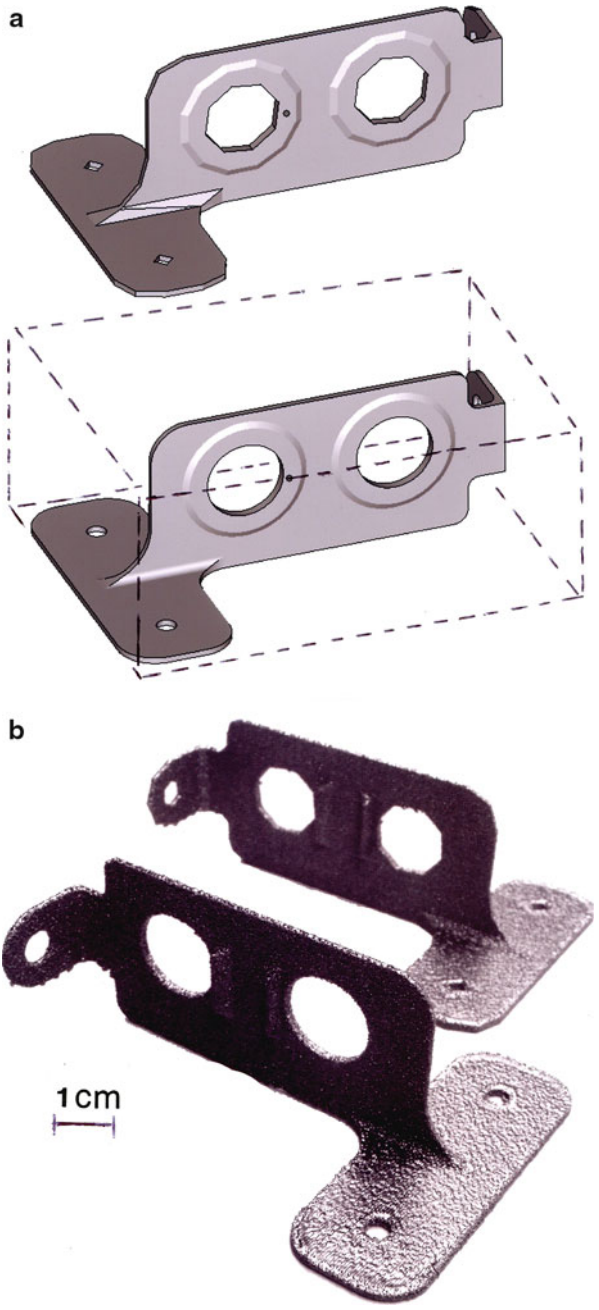


Fig. 1 Simple Ti-6Al-4V aircraft structural brackets fabricated by EBM. (**a**) STL (CAD) models. (**b**) EBM-fabricated components. The dotted rectangular volume in (**a**) represents the minimum raw material form for conventional production of the part

Fig. 2 Multi-movement gear component product fabricated by FDM. Part with build (ABS polymer) and support (waterworks) material is on the *left*. *Right* image shows working part with support material removed (Courtesy of Frank Medina, W.M. Keck Center for 3D Innovation, University of Texas at El Paso)

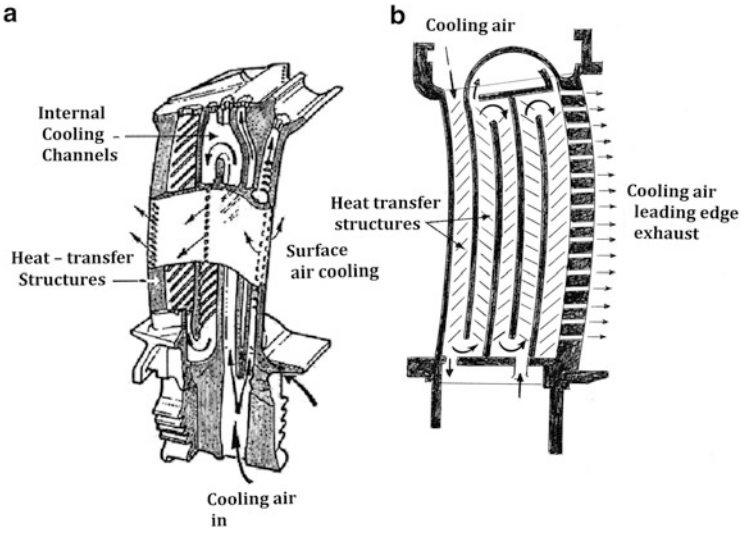


Fig. 3 Internal turbine blade cooling structures. (**a**) General conceptual sketch. (**b**) Complex cooling structure model adapted from 2010 patent by G. Liang (2010)

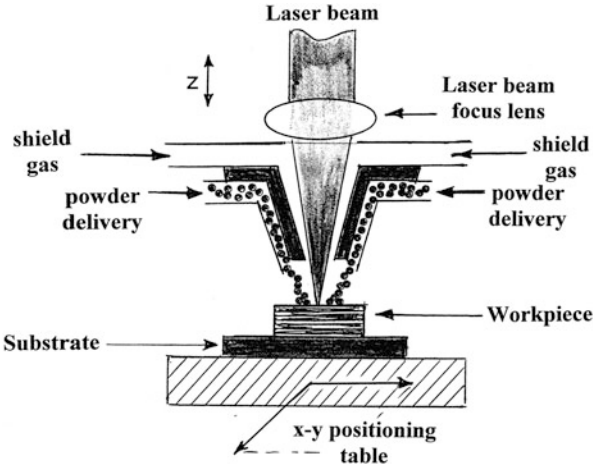


Fig. 4 Schematic diagram of the LENS process. Powder delivery occurs from a cylindrical-like tapered nozzle surrounding the focused laser beam. The process is often referred to as laser spraying

in particular has been the search for elemental additions to increase the high-temperature strength and suppress creep-rupture or increase their creep-rupture life (Moffura et al. 2012; see ▶ Figs. 3, ▶ 5, and ▶ 8 of chapter "Examples of Directional Crystal Structures: Gas Turbine Component Applications in Superalloys"). Another strategy for extending the high-temperature performance of more traditional high-temperature superalloy turbine blades involves more efficient cooling structures, particularly internal structures. Conventional casting strategies are unable to extend cooling channel complexities as illustrated in Fig. 4. However, small turbine blade complexity is easily accomplished by SLM or EBM fabrication, provided pre-alloyed powder precursors are available. In addition, the additively manufactured turbine blade products will be required to emulate the complex microstructures, such as precipitation optimization, which have been requisite materials performance issues for contemporary turbine blades. Murr et al. (2013) have recently shown that EBM fabrication of a contemporary Ni-base superalloy composition can produce optimized performance microstructures, virtually eliminating complex, postproduction heat treatments. ▶ Figure 18 of chapter "Examples of Directional Crystal Structures: Gas Turbine Component Applications in Superalloys" shows optimized γ' precipitation (nanoparticle size and shape) as well as surrounding matrix channel size (▶ Fig. 9 of chapter "Examples of Directional Crystal Structures: Gas Turbine Component Applications in Superalloys") for a Ni-base superalloy prototype fabricated by EBM. ▶ Figure 6b of chapter "Volume Defects: 3D Imperfections in Crystals" has already illustrated the details of this microstructure by TEM analysis.

There are many other additive manufacturing processes for metals. These include Laser Engineering Net Shaping (LENS) which incorporates a laser beam to melt blown powder on a surface using a configuration shown schematically in Fig. 4. This technique is often used to replace or replenish eroded surfaces and in effect "prints" new surfaces or replacement surfaces. In some respects, this technology is a modification of the more traditional weld replacement or cladding technologies which have been used since the mid-twentieth century to build up

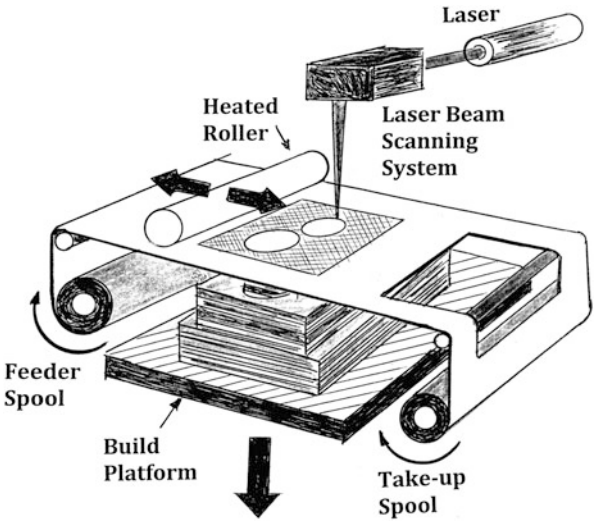


Fig. 5 Laminated object manufacturing (LOM) process schematic

eroded surfaces on large components such as large impellers in dredging machines and the like. In contrast, SLM is able to make far more geometries than LENS.

Laminated Object Manufacturing (*LOM*) *and Ultrasonic Consolidation* (UC) are layer manufacturing technologies which utilize material sheets of various dimensions instead of precursor powders to build up a product or component. LOM (Jacobs 1992; Gibson et al. 2010; Park et al. 2000) uses adhesive foils or sheets as basic building units which are hot pressed or roll-bonded. The UC process uses metal or alloy sheets where sheet bond is characterized microstructurally by micrograin migration, or the mechanical mixing of recrystallized grains by solid-state flow to be described in more detail in Part XIII.

LOM is somewhat a lesser known 3D printing technology with some similarities to sheet rolling in lithographic processes, but employing laser cutting to define the build shape as illustrated schematically in Fig. 5. The process can create large parts relatively cheaply. As illustrated in Fig. 5, a sheet of laminated material is fed from a roll onto the build platform where it is melted by a heated roller to the substrate. A laser traces out the layer outline and excess material is wound up on a take up roll. The platform moves down and fresh material is rolled over it, repeating layer-by-layer fabrication. This process can work with paper or paper-related precursor sheets as well as polymeric or thin metal sheet materials. Adhesive-coated paper products or models fabricated by LOM can have wood-like characteristics and are usually sealed after fabrication to retard moisture absorption. Metal sheet fabrication has only been demonstrated with relatively low melting metals such as aluminum and some aluminum alloys.

Layer fabrication characteristic of UC processing uses ultrasonic energy to create a shallow bond between metal or alloy plates by solid-state flow of recrystallized grains as noted. Figure 6 illustrates a UC composite made from a laminate schedule of dissimilar metals (Janaki et al. 2007; Obielodan et al. 2010).

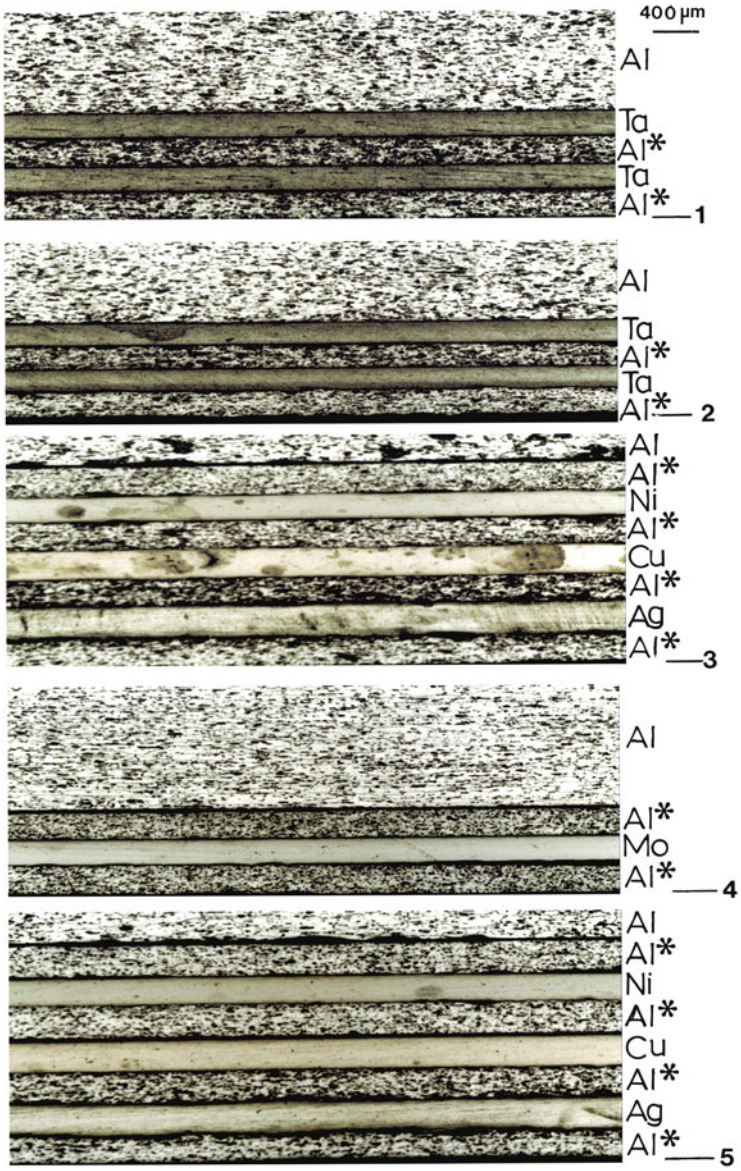


Fig. 6 Ultrasonic bonding of multi-metal laminate prototype products. Five different bonding arrays are shown. Al* denotes aluminum alloy 6061

The potentials of both LOM and UC technologies have yet to be fully exploited. One recent, ultrasonic application related to additive or 3D printing processes involves the ultrasonic embedding of wires connecting components in embedded electronic devices (Espalin et al. 2014). Embedded copper wires having diameters

averaging 20 μm have a distinct advantage over 3D printed interconnects, especially in terms of reliability.

It is clear that 3D printing and the more general technologies representing additive (layer) manufacturing can form a wide variety of hybrid manufacturing systems, even integrating aspects of more traditional subtractive manufacturing with the vast varieties of additive manufacturing technologies. It might also be appreciated that additive manufacturing systems can be employed in concepts involving mobile factories, fabricating replacement parts or one-of-a-kind parts in remote locations or in nontraditional locations, even "manufacturing malls" with 3D printers or variations producing novel, customer-specific, or, in a medical context, patient-specific products. Even producing novel products by digital scanning objects can be replicated and selectively modified in CAD data files to build exact or modified replicas or appropriately modified copies.

3D printing CAD has been referred to as "The Desktop Factory" (Binns 2007) and more recently as "A Factory in the Classroom" (Lipson and Kurman 2013) as an indication of the pervasiveness of these technologies. The concept of *modular manufacturing* can also integrate 3D printing or related additive manufacturing technologies with bonding or joining strategies such as friction welding or friction-stir welding and processing to be described in Part XIII. A modular manufacturing factory utilizing an EBM or SLM system, a CNC multiaxis machine for product finishing along with a friction-stir welding (FSW) system for novel joining, and a high-definition finishing area, such as enclosed electroplating or electroetching, could occupy an area of 2,500 ft^2, including CAD workstations. Characteristic of all 3D printing and AM processing, these manufacturing modules would be energy efficient, material efficient, and environmentally efficient, capable of providing either multiple product lots or one-of-a-kind products, including unique tooling such as special molds or injection molding components and process structures. All of this will lead to hundreds of thousands of small and medium-sized manufacturing businesses, new materials and materials applications, cheaper robotic and software concepts, an abundance of 3D printers and additive manufacturing systems, and novel, customer-specific online services.

References

Binns C (2007) The desktop factory in *Popular Sci*. Bonnier, New York, pp 42–44

Espalin D, Muse DW, Medina F, MacDonald E, Wicker RB (2014) 3D printed electronics for rapidly deployable satellite applications. Int J Adv Manuf Technol 72(5–8):963–978

Gibson I, Rosen DW, Stucker B (2010) Additive manufacturing technologies: Rapid prototyping to digital manufacturing. Springer, New York

Jacobs PF (1992) Rapid prototyping and manufacturing. Soc Manuf Engrs, Dearborn, MI.

Janaki Ram GD, Robinson C, Yang Y, Stucker B (2007) Use of ultrasonic consolidation for fabrication of multi-material structures. Rapid Prototyp J 13:226–235

Liang G (2010) Large chord turbine van with serpentine flow cooling circuit. US patent no 07785072, issue date 31 Aug 2010

Lipson H, Kurman M (2013) Fabricated: the new world of 3D printing. Wiley, New York

Moffura A, Finnis MW, Reed C (2012) On the possibility of rhenium clustering in nickel-based superalloys. Acta Mater 60:2866–2872

Murr LE, Martinez E, Pan XM, Gaytan SM, Castro JA, Terrazas CA, Medina F, Wicker RB, Abbott DH (2013) Microstructures of Rene 142 nickel-base superalloy fabricated by electron beam melting. Acta Mater 61:4289–4296

Obielodan JO, Ceylan A, Murr LE, Stucker B (2010) Multi-material bonding in ultrasonic consolidation. Rapid Prototyp J 16(3):180–188

Park J, Tari MJ, Hahn HT (2000) Characterization of the laminated object manufacturing (LOM) process. Rapid Prototyp J 61(1):36–50

Part XII

Nanomaterials: Structure, Properties, Processing and Performance

Serendipitous Nanotechnology in Antiquity

Contents

Abstract

Like Mayan blue, involving indigo dye trapped in palygorskite nanofibers (chapter "▶ Examples of Materials Science and Engineering in Antiquity"), Egyptian blue and Chinese (or Han) blue and purple represent nano-pigments with rich roles in Egyptian and Chinese antiquity. Kaolin nanosilicate sheets also contributed to Chinese porcelain in antiquity. While these examples represent nano-particulate materials, the role of carbon and carbide nanoparticles in bulk iron matrices and related accidental nanotechnology in antiquity, manifested in the development of Damascus and related patterned swords and daggers, represents the application of the materials science and engineering paradigm in antiquity: structure–properties–processing–performance interactions and relationships. These issues are presented in this chapter which, along with other developments in the many parts of this handbook, attest to the evolution of nanomaterials and nanotechnologies over many millennia.

Introduction

Nanomaterials, as implicit in retrospect, compose the smallest length-scale component of essentially all matter systems, both living and nonliving. As noted in Part I, especially chapter "▶ Examples of Materials Science and Engineering

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_44

Fig. 1 Examples of Egyptian tomb wall hieroglyphics and paintings using Egyptian *blue* pigment (Courtesy of Sandra Vannini/CORBIS)

in Antiquity," there have been many examples of nanomaterials and serendipitous or accidental nanotechnologists and nanotechnologies in antiquity.

It has already been demonstrated that color, especially structural color, has its origins in the nanoscale, if not the hundreds of nanometers. But color has had, and of course continues to have, an important role in the development of civilizations, especially in aesthetic stimulation as reflected in various art forms, and the application of dyes or pigments in wall paintings. These dyes or pigments were derived primarily from plants or soils (mineral mixtures) where the color blue was generally absent, except for copper-bearing blue minerals in segregated deposits containing azurite, malachite, lapis lazuli, and the like. It is possible that because of this lack of blue pigmentation, blue became held in high regard in the ancient world, long before the development of Mayan blue as illustrated in ▶ Fig. 9, chapter "Examples of Materials Science and Engineering in Antiquity".

Egyptian Blue Nanosilicate Pigment

Sometime around 2575 BC, the Egyptians developed Egyptian blue, regarded as the first synthetic pigment and purported to emulate the blue color of copper-bearing minerals. It is generally thought that Egyptian blue, found in tomb wall paintings as illustrated in Fig. 1, statues and other structures, and papyri, was manufactured by grinding sand (or quartz: SiO_2), limestone ($CaCO_3$), and copper minerals such as azurite, malachite, or tenorite (CuO) and firing at ~900 °C to produce a layer copper silicate, cuprorivaite ($Ca\ Cu\ Si_4\ O_{10}$): CaO: CuO: $(SiO_4)_4$;

a gillespite-structured phase having a complex, tetragonal crystal structure with a = 7.30 Å, c = 15.10 Å; and Space Group: P4/ncc. The actual tetragonal crystal unit cell is composed of 4[Ca Cu Si_4 O_{10}] and contains four SiO_4 tetrahedra linked together by square-planar coordinated copper (Kendrick et al. 2007), the source of the blue color. Very recently, Kendrick et al. (2007) discovered that the Ca Cu Si_4 O_{10} tetragonal crystals composing Egyptian blue can be delaminated in hot water (80 °C) forming a colloidal dispersion of < 10-nm-thick nanosheets and > 10-nm-thick nanoplatelets of Ca Cu Si_4 O_{10}. These authors demonstrated that this colloidal dispersion of Ca Cu Si_4 O_{10} (Egyptian blue) mixed with N-methylpyrrolidone could print patterns on a glass substrate using an ink-jet printer. In addition, Egyptian blue pigment was also shown to exhibit strong near-IR emission properties which suggest a new class of 2D nanomaterials applicable in near-IR-based biomedical imaging, IR light-emitting devices, and security ink formulations (Johnson-McDaniel et al. 2013).

Chinese Blue and Purple Nano-Pigments

Some two millennia after Egyptian blue was developed, the Chinese produced a blue pigment phase similar to Egyptian blue, but with Ba replacing the Ca: Ba Cu Si_4 O_{10}, having the same tetragonal crystal structure and Space Group, with lattice parameters a = 7.44 Å, c = 16.14 Å (Kendrick et al. 2007). This Chinese blue (also called *Han blue*) was originally produced near Xi'an, China, in connection with efforts to create glass-emulating jade, which in early Chinese culture symbolized immortality. Han blue (or the mineral *effenbergerite*) was manufactured similar to Egyptian blue, with barite ($BaSO_4$) substituting for limestone ($CaCO_3$) or lime (CaO). However, it is not generally thought that the production of Egyptian blue influenced Chinese blue production.

Chinese blue, however, was observed to be somewhat unstable, and it could produce a more stable, deep blue or purple pigment called Chinese purple (or *Han purple*):

$$3\ BaCuSi_2O_6 \rightarrow BaCuSi_4O_{10} + 2\ BaSiO_3 + 2\ CuO$$
$$(\text{Han purple}) \rightarrow (\text{Han blue})$$

In addition to the coloring of ceramic products, especially colloidal (nano) clay products such as Chinese clay (or kaolinite), also a layer silicate as illustrated in Fig. 2, Han blue and purple pigments were also used in coloring the terracotta warriors constructed in connection with the burial site of Qin Shi Huang, the first Chinese emperor, in Xi'an, Shaanxi Province, China. Ying Zheng, who took the Chinese throne in 246 BC at the age of 13, also took the name Qin Shi Huang Di – the First Emperor of Qin, after unifying seven warring kingdoms into a single empire called China. The emperor also developed extensive road systems and canals connecting provinces and standardized the units of measure, currency, and the Chinese script. Qin Shi Huang also began the construction of a precursor

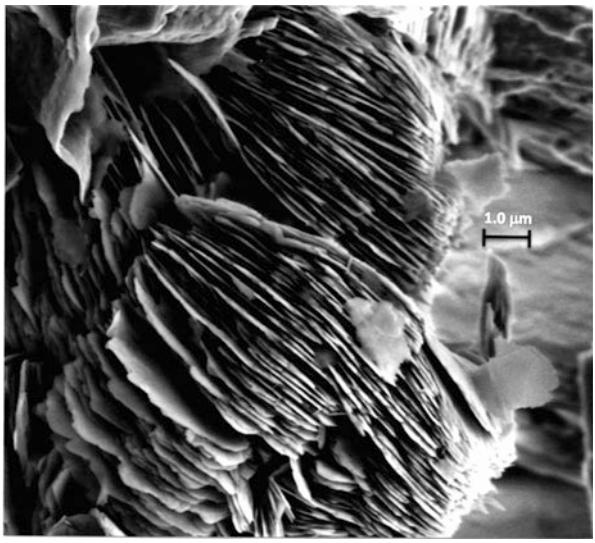


Fig. 2 Kaolin (kaolinite: $Al_2Si_2O_5$ $(OH)_4$) nano-sheet silicate observed in the SEM

to the Great Wall, which, when completed in all of its various phases, has been estimated to have exceeded a total length of 13,000 miles (21,666 km).

According to historians such as Sima Qian (145–90 BC), work began on the tomb (or mausoleum) of Emperor Qin shortly after he ascended to the throne (in 246 BC). In his historical account, *Shiji*, Qian wrote that the burial complex, including a tomb that dwarfed the pyramids of Egypt (▶ Fig. 1 in chapter "Examples of Materials Science and Engineering in Antiquity"), was the size of a city and may have involved as many as 700,000 workers. The tomb, constructed at the foot of Mount Li (a favored location because of its famed jade mines) as an earthen pyramid, was said to contain replicas of palaces and the earthly realm of the Emperor, 100 rivers simulated with flowing mercury, and a copper ceiling decorated with gem stones representing heavenly bodies. Interestingly, no historical accounts recorded the terracotta warriors and the collection of terracotta sculptures depicting the armies of Qin Shi Huang, which were discovered by three farmers digging a water well in a location 1.6 Km east of the Emperor's tomb, on March 29, 1974. Since this time, more than 2,000 life-size soldiers measuring ~1.8 m in height, each weighing about 150 kg, in three subterranean pits, sculpted individually in clay, each with a different likeness, and lavishly painted and lacquered, have been uncovered along with horses, chariots, and weapons. The terracotta figures (warriors and horses) have mostly lost their coloring which is represented as shown in Fig. 3 along with examples of the warrior columns and horses, as well as a photograph of the tomb mound. The tomb mound represents the pyramid, covered over the millennia by up to 5 m of sandy soil. Originally, the tomb measured 115 m in height and had a base covering more than 121,000 m^2. The blue rendering in the insert in Fig. 3c represents the Han blue pigment, while it has been determined from more recent and more carefully preserved discoveries that many figures, especially

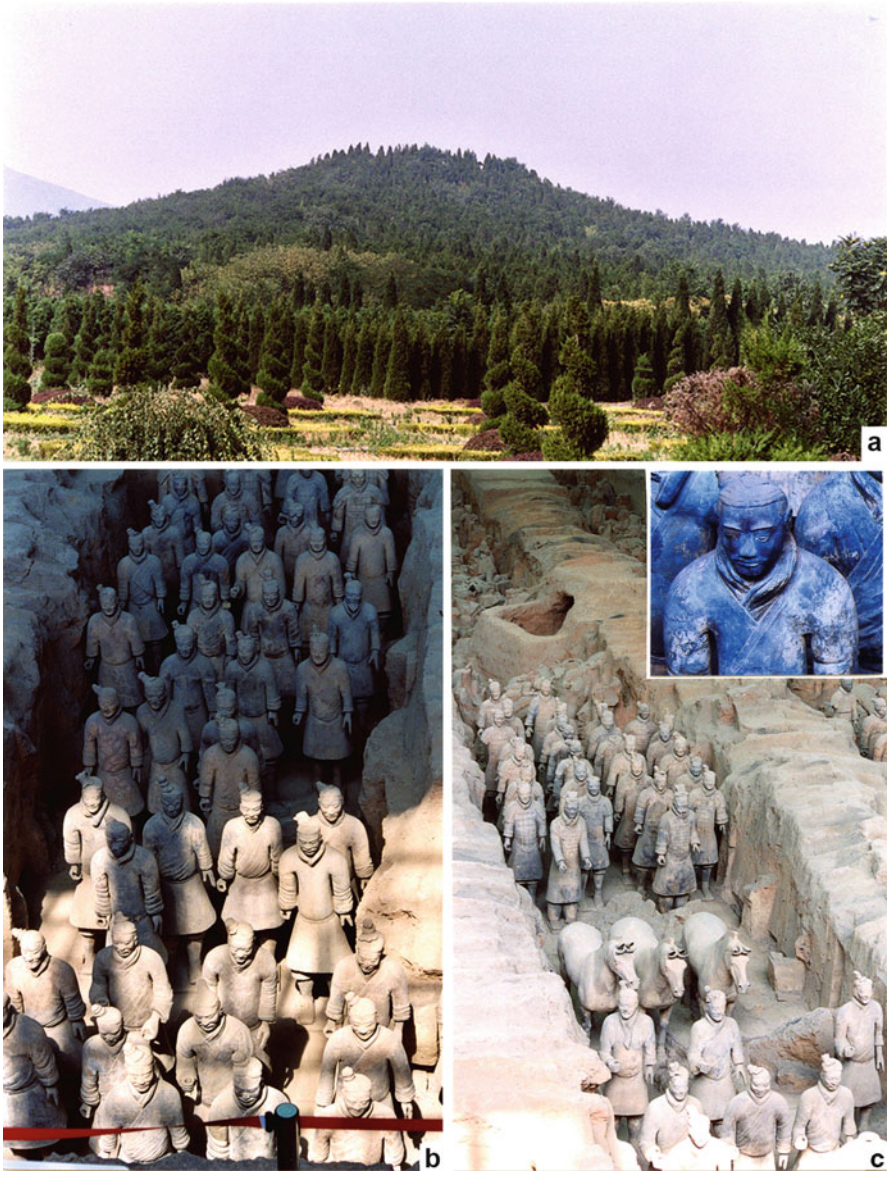


Fig. 3 Qin Shi Huang tomb, Xi'an, China (**a**), and terracotta warrior pits (**b**) and (**c**). The insert in (**c**) shows a blue-painted (*Han blue*) warrior recently uncovered. Clay figures were arranged four abreast as shown

ranking soldiers, were colored with Han purple, in particular their trousers. By the end of 2012, an estimated 6,000 terracotta soldiers remained to be uncovered, but they, like the elaborate tomb of the Emperor, remained unexplored because of concerns for degradation and especially the rapid deterioration of tomb artifacts on

exposure to the air. In addition, the descriptions of mercury rivers seem credible in light of thousands of soil samples taken from throughout the tomb mound (Fig. 3a) over the past decade which invariably show mercury levels by orders of magnitude above background.

Weapons, including bronze and steel swords (some coated with chromium oxide which made them corrosion resistant) spears, battle-axes, crossbows, and arrow-heads, were found in the warrior pits illustrated in Fig. 3b and c. These represented iron and bronze (Cu and Sn alloys), some containing other metallic elements such as Ni, Co, and Mg. Wooden shafts and other components were destroyed by bacteria and other organisms over the millennia. However, it was known that in 206 BC, 4 years after the death of Emperor Qin, the burial vaults containing the terracotta soldiers were vandalized by peasants who stole many of the weapons and used them in a rebellion against Qin Shi Huang's descendants. Supporting roof structures set on fire also burned paint and glaze from the clay figures. Many other bronze artifacts were uncovered along with items made of gold and silver. Many jade carvings have also been recovered. Similar artifacts are indicated in the historical records to be in the main tomb as well (Luo 1993).

Just as important nanomaterial properties have recently been explored for Egyptian blue (Ca Cu Si_4 O_{10}) as discussed by Johnson-McDaniel et al. (2013), similar, contemporary nanomaterials phenomena have been observed in Chinese (Han) purple (Ba Cu Si_2 O_6) by Kramer et al. (2012). In particular, the nanolayer structure gives rise to a two-dimensional Bose–Einstein condensed phase which effects local magnetization. Consequently, Egyptian blue and Han purple pigments developed and utilized by the ancients millennia ago are being rediscovered for potential applications in contemporary nanotechnology.

Iron and Steel in Antiquity: The Role of Nanocarbon

As noted in chapter "► A Brief History of Metals," the Iron Age is usually extended to ~1500 BC based on various historical findings. However, the translation of Agricola's *De Re Metallica* by Herbert Hoover and his wife, Lou, (Hoover and Hoover 1950) speculates a period extending as long as 5000 BC. Iron daggers were found on Tutankhamun's mummy which is known to be from 1350 BC (Sherby 1999). A thin, laminated iron plate was also discovered in the Great Pyramid at Giza (Khufu's pyramid shown in ► Fig. 1 in chapter "Examples of Materials Science and Engineering in Antiquity") in 1837 which was analyzed by Gayar and Jones (1989) nearly a century and a half later. As noted above, steel weapons were found among the terracotta warriors dating to the time of Emperor Qin circa 250 BC. Cast iron was made in China as early as 200 BC. Cast iron with a carbon content of 2–4 % requires higher furnace temperatures and more reducing conditions than the bloomery iron process. Hittite records (1700–1200 BC) suggest they were aware of both meteoric iron and smelted iron, and iron seems to have been used in India from about 1500 BC. In fact, the earliest large iron forging (99.7 % Fe) in the world is the famous iron pillar at New Delhi dated by inscription to the Gupta period (circa

Fig. 4 Corrosion-resistant Delhi iron pillar (Delhi, India)

400 AD). This pillar, shown in Fig. 4, is over 7 m in height and weighs roughly 6 t. The fame of this pillar lies in the fact that it has appeared to be "rustless" over more than 1,600 years. However, Balasubramaniam (2000) has identified a protective passive layer composed of iron hydrogen phosphate hydrate ($FePO_4 \cdot H_3PO_4 - 4\ H_2O$) in a crystalline form, in addition to α-, γ-, and δ-FeOOH and magnetite (Fe_3O_4) all in amorphous form. Slag particles also play a role in this matrix. The low porosity of the crystalline phosphate phase lowers the rate of corrosion. So while a thin "rust" layer exists, it allows only very slow corrosion of the underlying iron pillar and gives the appearance of being rustless.

Originally, pure iron (or carbon-free iron) when heated to the maximum temperature of ancient furnaces (~1,150 °C) did not melt, since its melting temperature is 1,539 °C. However, at ~1,100 °C, iron could be hammered or forged to purify and shape it without any carbon. Iron worked in this way is referred to as wrought iron. However, the Chinese, among others around 500 BC, learned that mixing carbon from charcoal into the iron bloom changed the difficulty in producing wrought iron (by consequently reducing the melting point). As noted previously, the Chinese produced cast iron (iron containing 2–4 % carbon) in small blast furnaces, which were actually precursors to modern blast furnaces.

In the initial smelting of iron, bloomery iron is the first product, a solid, spongy mixture of iron and slag. By hammering (or *smithing*), excess slag is forced out. This process alternates annealing at ~1,000 °C and hammering. As carbon was added, low-carbon steel emerged as high-carbon steel and cast iron. Early products involved plows, hammers, axes, spear heads, and swords. Swords and sword making are among the more interesting examples of iron and steel products in

antiquity, especially the evolution or etymology of *Damascus swords* and related patterned swords.

Before exploring the development of swords, especially novel steel swords in antiquity, it would seem useful to examine the iron–carbon (binary) phase diagram, at least the most important part of the diagram which can act as a microstructural–temperature–composition overview, especially as the microstructures relate to nanocomponents or nanostructures. Historically, European metallurgists had only begun to understand binary alloy systems such as iron and carbon by the late nineteenth century. Two of the most influential European metallurgists contributing to a fundamental understanding of iron and steel were Henry Clifton Sorby (1826–1908), an English microscopist, geologist, metallurgist, and perhaps the founder of metallurgy, and Floris Osmond (1849–1912), a French metallurgist and metallographer. Sorby, after investigating the microscopical structure of rock and minerals, applied microscopical analysis to investigating acid-etched structures and phases in iron and steel in 1863. These included cementite (Fe_3C eutectic) and pearlite (lamellar α-iron (ferrite) and cementite). It was Osmond in 1885 who actually identified Fe_3C as "ciment," French for binder or cement. He named pearlite "sorbite" in honor of Sorby, but the term pearlite, originating from the pearl-like luster and iridescence it produces, persisted. Osmond, in 1890 also recognized the α, β, and γ modifications or phases of iron in the context of the iron–carbon diagram which, as illustrated in Fig. 5, is essentially a graph showing the relationship between these phases, the iron–carbon composition, and the temperature. Sir William Roberts–Austen (for whom austenite or γ-iron was named) established the iron–carbon phase diagram in 1989, which in fact was the first establishment of an alloy phase diagram. Final touches to this diagram were made by Korato Honda, a pioneering Japanese steel metallurgist, in 1920.

It can be noted that in the Fe–C phase diagram portion of Fig. 5, the three prime phase constituents are defined in the key box below the figure. The range of steel and cast iron are indicated along with the hypoeutectoid and hypereutectoid regimes in steel. The delta (δ) phase or *δ*-Fe is shown shaded in the upper left portion of Fig. 5, while Fig. 6 shows a rare TEM image recently obtained for thin, irregular δ-Fe nanoplates coincident with either {110} or {100} planes in α-Fe (see Murr et al. (2013)). In addition, Fig. 7a illustrates regular, lamellar pearlite in gray iron in contrast to nodular or spheroidal pearlite (spherical Fe_3C inclusions) shown in Fig. 7b. Figure 7c illustrates graphite flakes in cast iron. In contrast to the pearlite in Figs. 7a and 8 shows the structure of a freshwater pearl surface which illustrates the lamellar aragonite ($Ca\,CO_3$) plates in contrast to the lamellar pearlite (α-Fe + Fe_3C). It can be observed in Fig. 7a that the Fe_3C (cementite) plates have a thickness of roughly 200–400 nm, while the aragonite tile plates in Fig. 8 have a thickness of roughly 300 to 400 nm. Similarly, the cementite (Fe_3C) nodules in Fig. 7b range from ∼200 nm to ∼2 μm. The graphite flakes in Fig. 7c range in thickness from < 1 μm to ∼6 μm. In addition, graphite nodules similar in appearance and dimensions to the pearlite nodules in Fig. 7b are also observed in cast irons.

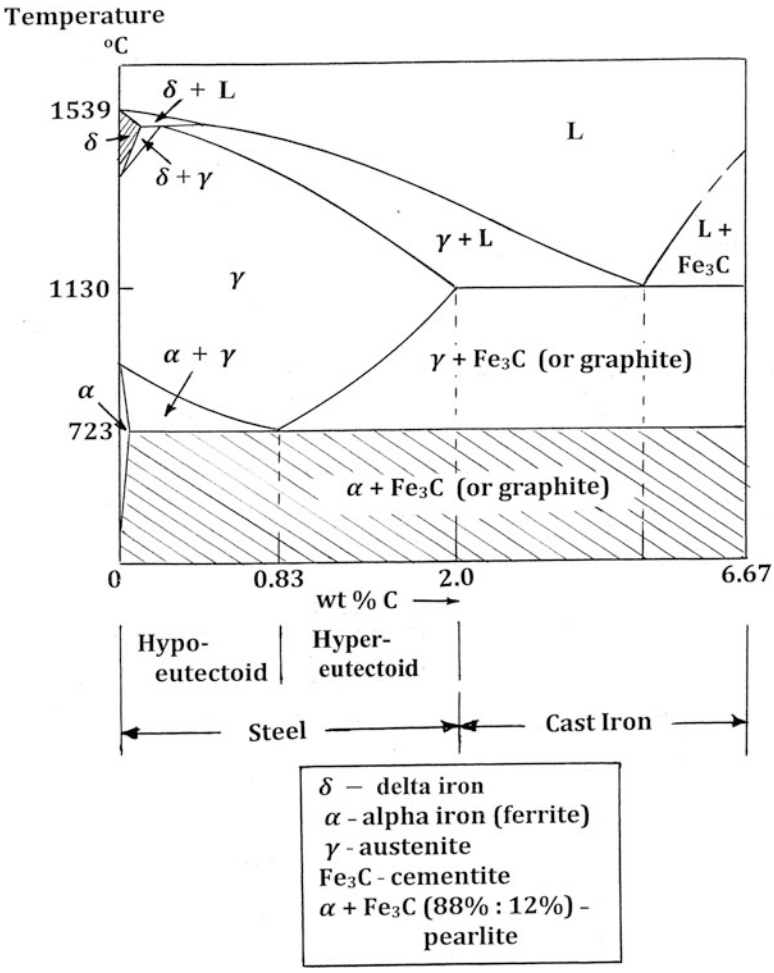


Fig. 5 Iron–carbon phase diagram

One of the more interesting features of iron and steel development in antiquity involves swords and sword making, particularly the Damascus and Samurai (Japanese) swords exhibiting unique blade patterns and superior properties, especially blade hardness and associated sharpness. The distinctive patterns of these swords have a similar microstructural origin characterized by bands of cementite (Fe_3C) particles (6–9 μm diameter) in an iron (α or γ) or pearlite matrix spaced ~30–70 μm (Verhoeven 2001).

There are two distinct metallurgical processes associated with these sword types. The Damascus swords, which along with similar swords made in India date to 300 BC, were made from hammered and annealed steel blocks or bricks containing ~1.5 % C or slightly more C. The hammer-anneal cycling by the bladesmith caused

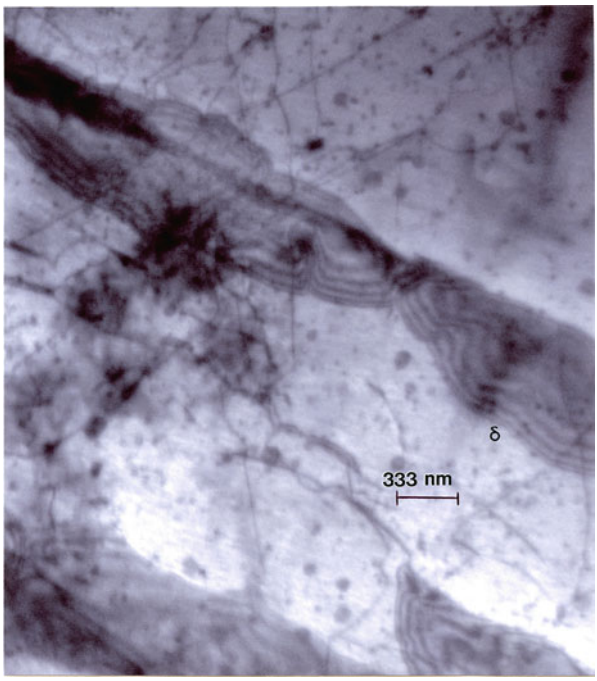


Fig. 6 TEM bright-field image of irregular δ-Fe platelets in α-Fe

cementite (Fe_3C) to break up into nodules resembling those shown in Fig. 7b, but forming in closely spaced sheets. These steel precursor blocks or ingots were mostly manufactured in the area of Hyderabad (formerly Gokonda), India. These ingots were used by bladesmiths in various parts of the world, particularly in the Damascus region to make swords and daggers with the unique surface patterns revealed by acid etching as illustrated in the simple patterns of Fig. 9. Figure 10 shows these patterns schematically, but in reverse contrast. The darker Fe_3C particle regions or bands shown in Fig. 10 become the white regions shown in Fig. 9 when etched, while the iron matrix appears dark. The Indian precursor ingots become known in the seventeenth century as wootz steel. Wootz is an anglicization of the word "ukka," which was the word for crucible steel encountered by European travelers in the Kannada-speaking regions of Karnataka and the Telugu-speaking region of Andhra Pradesh in India. Other variations of this steel ingot include the Persian "polad," Mongolian "bolat," Russian "bulat," and Turkish and Arab "fulad." The Damask (also Damast) or Damascene patterns derive largely from damas, which in the Arabic language means water. It was the interest by European metallurgists in the seventeenth century in trying to unravel the secrets of wootz steel and Damask patterning that led to the eventual development of the iron–carbon diagram summarized in Fig. 5.

The Japanese (Samurai) swords with unique surface patterns also derived from bands of Fe_3C in an iron matrix evolved from a different metallurgical process

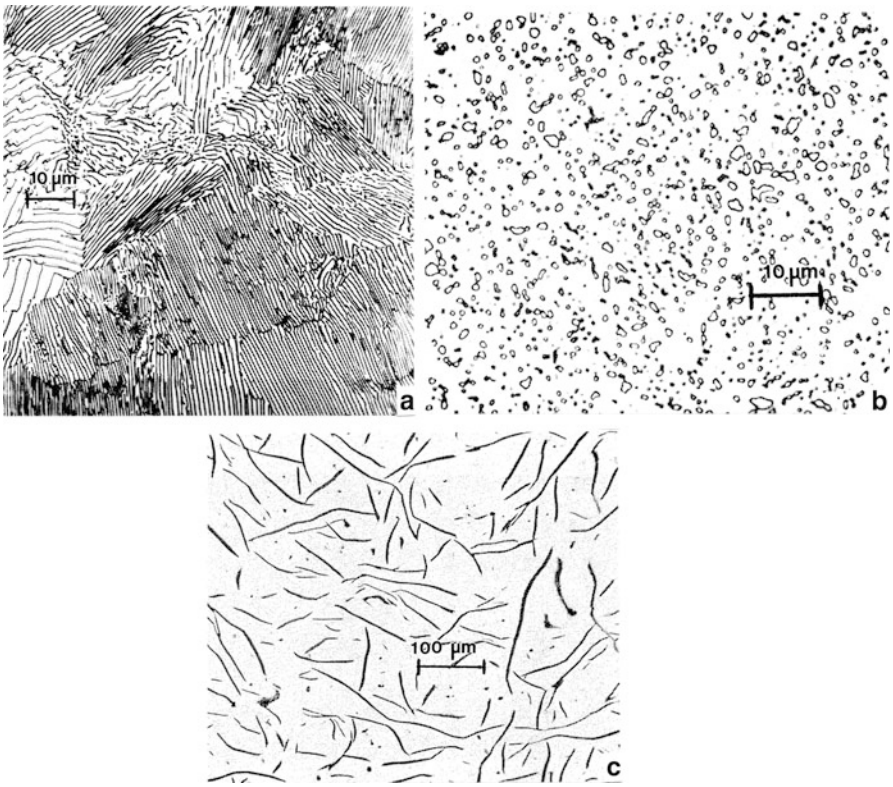


Fig. 7 Examples of carbon phases in iron. (**a**) Pearlite in gray iron. Lamellar colonies. (**b**) Modular (spheroidal) pearlite. (**c**) Graphite flakes in cast iron

involving laminates of steel or iron and steel, often referred to as laminated composites (Wadsworth and Lesuer 2000).

The laminated steel plate discovered in the Khufu Pyramid as discussed briefly above consisted of thin, alternating layers of low-carbon steel (~0.2 % C) and wrought iron (essentially pure iron). These laminated steels were made by *forge welding* or *pattern welding* where two or more alternating layers of lower and higher carbon steel are created by hammer welding them together. This process is also referred to as Damascene forging. Although some Damascus swords were also made by this process, the Japanese sword is probably the most famous. The Japanese swords postdate Chinese blades made by a similar laminating steel process referred to as "a hundred refinings" (circa 100 AD and onward). Japanese swords were first produced around 400–500 AD. The key to the acclaim of these swords was the intricate and time-consuming process of hammer-welding a bar of iron lengthwise to two bars of different grades (carbon contents) of steel. This triplex was then folded upon itself and hammered out again to its original length. This laminate was then coated with a clay-stem ash mixture, fired, folded again, and

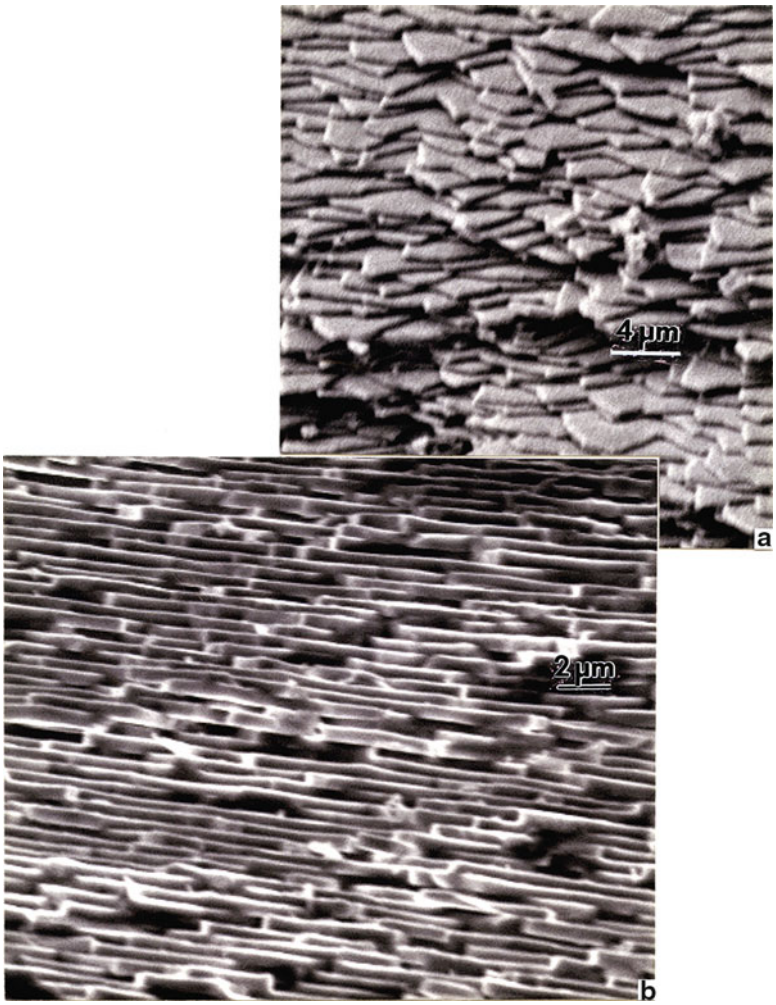


Fig. 8 Aragonite tile structure of the surface of freshwater pearls. (**a**) Near surface view. (**b**) Surface cross section

recoated with this mixture, and the process cycle is repeated 15 or 20 times to produce thousands of layers of steel. In some cases, the bladesmith would begin with three or four of the steel–iron–steel triplex, and when folded, hammered, and coated in multiple cycles, it could contain several million microscopic steel laminates (Robinson 1961; Kapp et al. 1987). These laminates were actually nanocomposites consisting of nanolayers of alternating or repeated sequencing of different carbon content nanolayers often $<$ 100-nm thick.

Tempering of blade edges of Damascus and Japanese swords involved both the art and science of bladesmithing. Japanese blades were so sharp that a square of silk

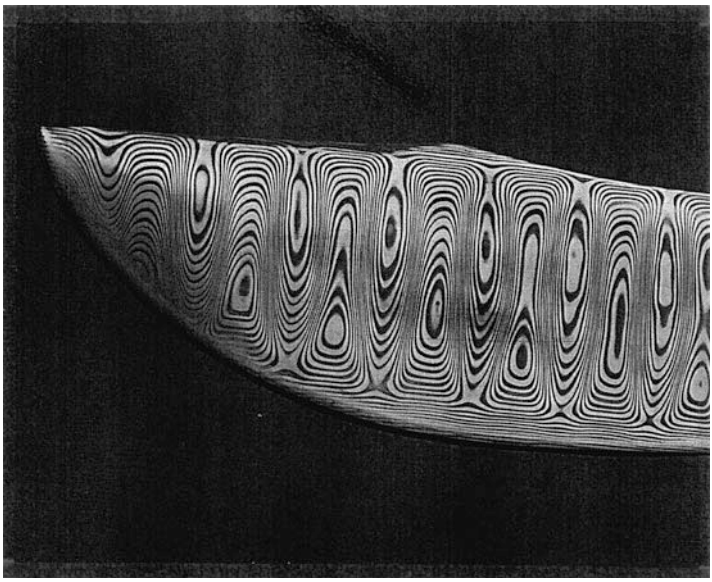


Fig. 9 Contemporary example of Damascene-patterned dagger

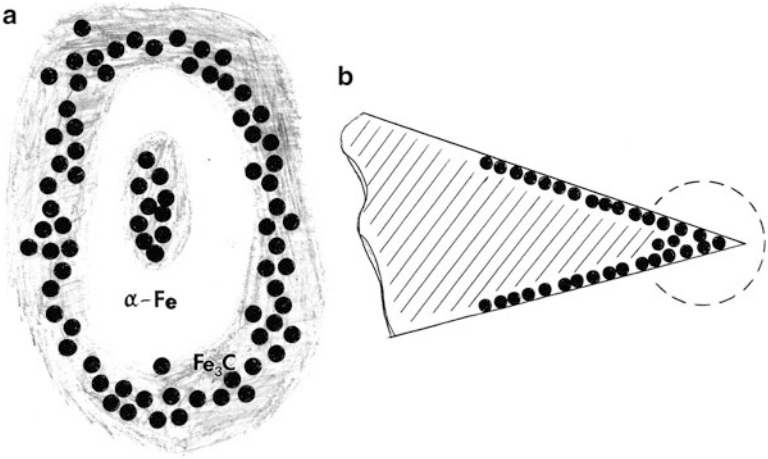


Fig. 10 (**a**) Schematic view of Fe_3C nodule bands in α-Fe matrix creating damascene etch patterns. Etching reverses the contrast. (**b**) Blade cross section illustrating concentration of carbides at tip contributing to high hardness and sharp edge

dropped onto an upturned blade would be sliced. Figure 10b shows the concentration of carbides in the blade edge, and this can schematically represent microstructural variations resulting by variations in blade heating and cooling schedules as well as bladesmithing cycles.

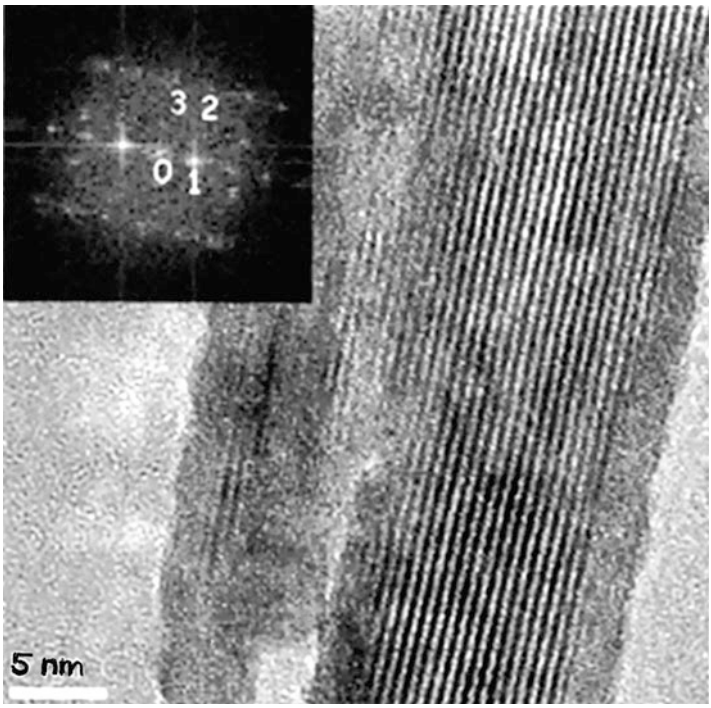


Fig. 11 High-resolution TEM image of cementite lattice fringes (0.63 nm) in remnant in Damascus sword section (From Reibold et al. (2009))

Reibold et al. (2009) recently found cementite nanowires and carbon nanotubes in a seventeenth-century (Damascus) sword using high-resolution TEM. In some cases, the cementite nanowires were encapsulated by carbon nanotubes. These observations suggest a complex relationship between cementite particles, impurity segregation, nanotube formation, cementite nanowire growth, and nanotube filling with cementite. In addition, this complex nanostructure may affect the mechanical properties of these swords. Figure 11 illustrates an embedded remnant characteristic of cementite lattice fringes observed in the TEM by Reibold et al. (2009). The other implication of observations such as the cementite nanowire in Fig. 11 is that the repeated hammer-annealing cycles in creating wootz steel and subsequent bladesmithing may also have refined considerable cementite to the nanoscale.

It is also clear that unusual strength resulted by ultrahigh carbon steel development. This feature has led to the reexamination of ultrahigh carbon steels in modern technology innovations. By 1975, J. Wadsworth and O. Sherby had observed that steels with carbon contents between 1 % and 2 % C could be both superplastic at warm temperatures and very strong and ductile at room temperature. Such steels had to be forged in a narrow temperature range of 850–650 °C in order to obtain the requisite fine grain size for superplastic behavior and the associated mechanical properties (Sherby 1999). In this respect, finely spheroidized cementite contributed

to yield strengths between 0.8 and 1.5 GPa, although more coarse cementite spheroids in ultrahigh carbon steels contribute to ductilities as high as 23 % tensile elongation.

An appropriate conclusion or summarizing phrase to this chapter might be an old Russian proverb: "The best of the new is often the long forgotten past."

References

Balasubramaniam R (2000) On the corrosion resistance of the Delhi iron pillar. Corros Sci 42:2013–2129

El Gayar ES, Jones MP (1989) Metallurgical investigation of iron plate found in 1837 in the Great Pyramid at Giza, Egypt. J Hist Metall Soc 23(2):75–83

Hoover HC, Hoover LH (trans) (1950) Georgius Agricola: *De Re Metallica*. Dover, New York

Johnson-McDaniel D, Barrett CA, Sharafi A, Salguero TT (2013) Nanoscience of an ancient pigment. J Am Chem Soc 135:1677–1679

Kapp L, Kapp H, Youhihara Y (1987) The craft of the Japanese sword. Kodansha, Tokyo

Kendrick E, Kirk CJ, Dana SE (2007) Structure and color properties in the Egyptian Blue Family, $M_{1-x}M^1{}_x\,CuSi_4O_{10}$ as a function of M, M^1 where M, M^1 = Ca, Sr and Ba. Dyes and Pigments 73:13–18

Krämer S, Laflorencie N, Stern R, Horvatic M, Benthier C, Nakamura H, Kimura T, Mila F (2012) Spatially resolved magnetization in Bose-Einstein condensed $BaCuSi_2O_6$: Evidence for imperfect frustration. Phys Rev B87:180405

Luo Z (1993) China's imperial tombs and mausoleums. Foreign Language Press, Beijing

Murr LE, Martinez E, Pan XM, Meng C, Yang J, Li SJ, Yang F, Xu Q, Hernandez J, Zhu WJ (2013) Microstructures and properties of solid and reticulated mesh components of pure iron fabricated by electron beam melting. J Mater Res Technol 2(4):376-385

Reibold M, Paufler P, Levin AA, Kochmann W, Pätzke N, Meyer DC (2009) Discovery of nanotubes in ancient Damascus steel. In: Physics and engineering of new materials. Springer proceedings in physics, vol 127. Springer, Heidelberg, pp 305–310

Robinson BW (1961) The art of the Japanese sword. Faber and Faber, London

Sherby OD (1999) Ultrahigh carbon steels, Damascus steels, and ancient blacksmiths. ISIJ Int 39(7):637–648

Verhoeven JD (2001) The mystery of Damascus blades. Sci Am 284:74–79

Wadsworth J, Lesuer DR (2000) Ancient and modern laminated composites – from the Great Pyramid of Gizeh to Y2K. Mater Charact 45(4–5):289–313

Classifications and Structures of Nanomaterials

Contents

Abstract

Nanomaterials are characterized by free nanoparticles of various nanodimensions: 1D, 2D, or 3D in classifying nanosheets such as graphene, 2D needles or filaments, and 3D particles, including aggregates. These geometries can be included in a contiguous solid as dispersed or precipitated phases or contiguous nanocrystals forming nanograin solids. These nanomaterials structures are generally observed directly using imaging techniques such as SEM, TEM, FIM, STM, or AFM, which are principal characterization tools. The distinctive features of nanomaterials especially nanocrystalline forms are the high surface-to-volume ratio and the dominant role of interfacial (and surface) energy. Nanostructured materials might be thought of as the interfacial state. These issues of characterization and classification of nanomaterials and nanostructures as well as interfacial phenomena in nanomaterials are treated in this chapter.

Introduction

It is of course aphoristic to state that nanomaterials, especially nanocrystalline materials, are not new. Examination of early investigations of age hardening in aluminum alloys in the early parts of the last century illustrates that submicron size

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_45

precipitates were responsible for this phenomenon (Wilm 1911; Merica et al. 1919). Moreover, this phenomenon represents another example of accidental nanotechnology which was later employed in revolutionizing the aluminum industry, and especially the application of aluminum alloys in the aircraft industry. The application of "nano" in describing ultrafine and submicron size and size-related phenomena in materials and material systems free particles, particles or phases as inclusions in solids, and ultrafine grain or phase solids did not become universally (globally) prominent until around 2000 (Siegel et al. 1999). Even after more than a decade, many books detailing aspects of nanomaterials, nanocrystalline materials, and nanotechnologies emphasize nanoparticulate phenomena even though Siegel (1993) in earlier work classified nanostructured materials into essentially four dimensional categories: zero-dimensional (0D) nanoparticles or nanoclusters, one-dimensional (1D) layer or multilayer nanostructures, two dimensional (2D) nanograined layers, and three dimensional (3D), equiaxed bulk solids. Gleiter (2000) also classified nanomaterials according to their chemical composition and crystallite shapes composing solid structures or nanostructures.

Characterization of Nanomaterials

Characterization of nanomaterials is here described as observing nanomaterials and their structures or nanostructures. Such observations have already been illustrated throughout this handbook for nanoparticles either in a free form or some nanophase in a solid matrix, such as dispersed or precipitated nanophases. The reader can peruse previous parts and chapters for many such examples. In spite of these many examples, it seems somewhat useful to examine a few available techniques for observing the nanoscale and to provide some additional and coherent as well as representative examples which will also begin to establish nanomaterials classifications, such as they are.

Electron microscopy, both scanning and transmission electron microscopy (SEM and TEM, respectively), are perhaps the most available and often useful characterization tools because nanomaterials can be readily prepared and examined directly, at resolutions of $\sim$1–10 nm in the SEM and $<$1 nm in the TEM, where in high-resolution instruments atomic lattice structures can be observed either at single atom resolutions or resolution of crystal lattice planes (Murr 1991). Figure 1 illustrates these nanostructural observations in the TEM. While the SEM is able to observe nanoparticle geometries, it can also examine chemical compositions using energy-dispersive (X-ray) spectrometry. This capability is also somewhat standard in TEM, which additionally affords a selected-area electron diffraction (SAED) pattern which can provide specific crystallographic information. These features are illustrated in Fig. 2 for small TiO_2 (tetragonal rutile) particles collected from the outdoor air and presumably resulting from the erosion of contemporary paints, which utilize nanoparticulate TiO_2 as the white base for paint. The TEM images in Fig. 2 illustrate diffraction contrast fringes around the particles which are observed to have crystal facets. The diffraction patterns illustrate the single-crystal nature for

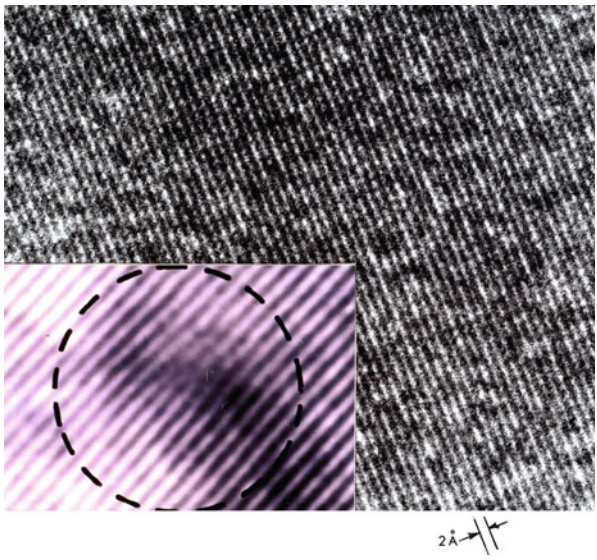


Fig. 1 High-resolution TEM images of atomic-level images. Large image shows near atomic-level image in Au. Insert shows dislocation (lattice) image in Pd. Both are fcc metals. Lattice image in Au film is (200) planes spaced 2 Å

the TiO_2 nanoparticles, while the energy-dispersive (X-ray) spectrum in Fig. 2c attests to the Ti and O composition. Figure 3 shows a number of different commercially manufactured nanoparticulate materials observed in the TEM along with their corresponding SAED patterns. Such nanoparticles are often electrostatically aggregated. Note the variations in sizes and shapes, which include spherical and sometimes faceted Fe_2O_3 nanoparticles in Fig. 3a, irregular nanocrystal rods of TiO_2 (tetragonal anatase) in Fig. 3b, NiO nanocrystal plates in Fig. 3c, and faceted (cuboctahedral) Al_2O_3 nanoparticles in Fig. 3d.

Figure 4a and c show TEM images of carbon nanotubes mixed with fullerene nanoparticle aggregates, while Fig. 5b shows an SEM image of nanosoot aggregates, which have been described in detail in Part III, chapter "► Chemical Forces: Nanoparticles." The TEM image of carbon nanotube and fullerene aggregates in Fig. 5a includes a corresponding SAED pattern which indicates some degree of crystallinity of the aggregated nanoparticles.

In contrast to the crystalline nanoparticle images shown in Figs. 2, 3, and 4 and observed by SEM and TEM, Fig. 5a and b show bacteria attached to mineral surfaces observed by SEM, while Fig. 5c and d show TEM image cross-sections of these bacteria using a selective stain to provide mass-thickness contrast (Murr 1991). The arrows in Fig. 5c illustrate the bacterial wall structure to consist of three molecular layers having a combined thickness representing the cell wall of ~35 nm.

Figure 6 shows tetragonal tin nanoparticles grown from the vapor onto NaCl (001) single-crystal substrates, and then extracted in their exact geometrical positions by vapor depositing a carbon film over the particles and removing the film with attached nanoparticles by dissolving in water. Figure 6a shows a bright-field image with an SAED pattern insert. The spots on the inner Sn ring indicate the particles formed epitaxially on the (001) NaCl surface, and crystallized in the c-axis

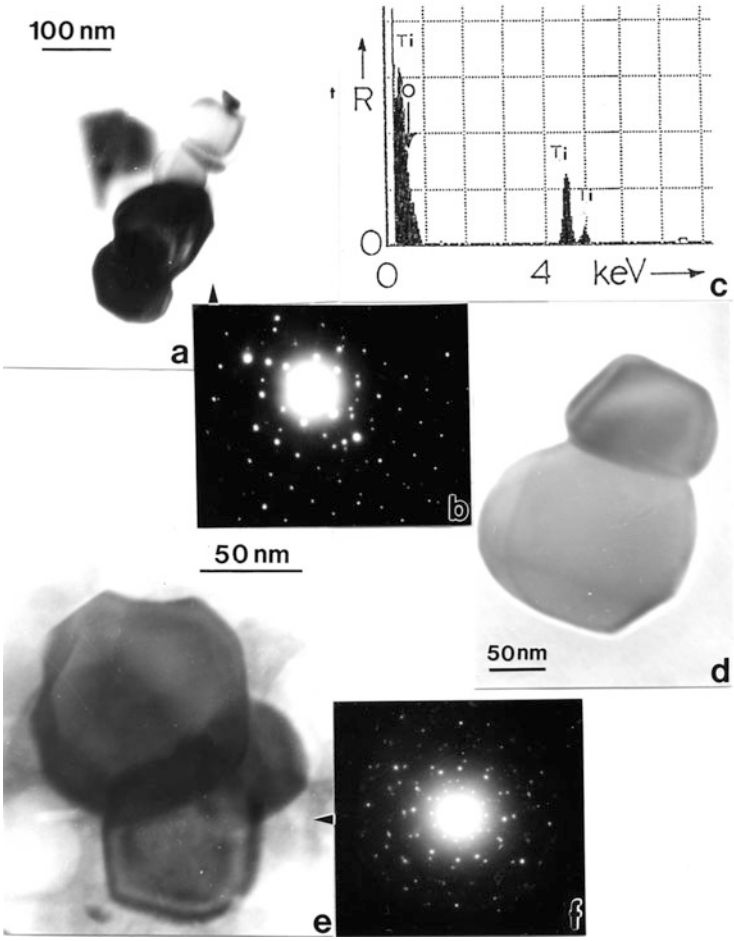


Fig. 2 TiO_2 nanoparticle aggregates (sintered nanoparticles) captured from the air. (**a**), (**b**), and (**c**) are representative of five particles. (**a**) TEM image. (**b**) SAED pattern of (**a**). (**c**) Energy-dispersive (X-ray) spectrum showing Ti and O. (**d**) TEM of two sintered particles. (**e**) and (**f**) show small TiO_2 cluster and SAED pattern

direction, with the basal (001) tetragonal unit cell parallel and epitaxial with the NaCl (001) surface plane. Figure 6b shows a dark-field image of Fig. 6a using the circled reflection (diffraction spot) in the SAED pattern insert. A bimodal distribution of vapor-grown Sn single-crystal particles is apparent, having mean particle sizes (diameters) of ~30 nm and ~400 nm.

In contrast to Fig. 6, Fig. 7 shows In nanocrystal growth on NaCl (001) substrates. In Fig. 7a, In nuclei have grown together to form a contiguous thin film composed of In crystal grains ranging in size from <50 nm to >500 nm, having a film thickness of ~200 nm. Correspondingly, Fig. 7b shows the formation of the contiguous (solid) film in Fig. 7a by converging (coalescing) tetragonal (001)

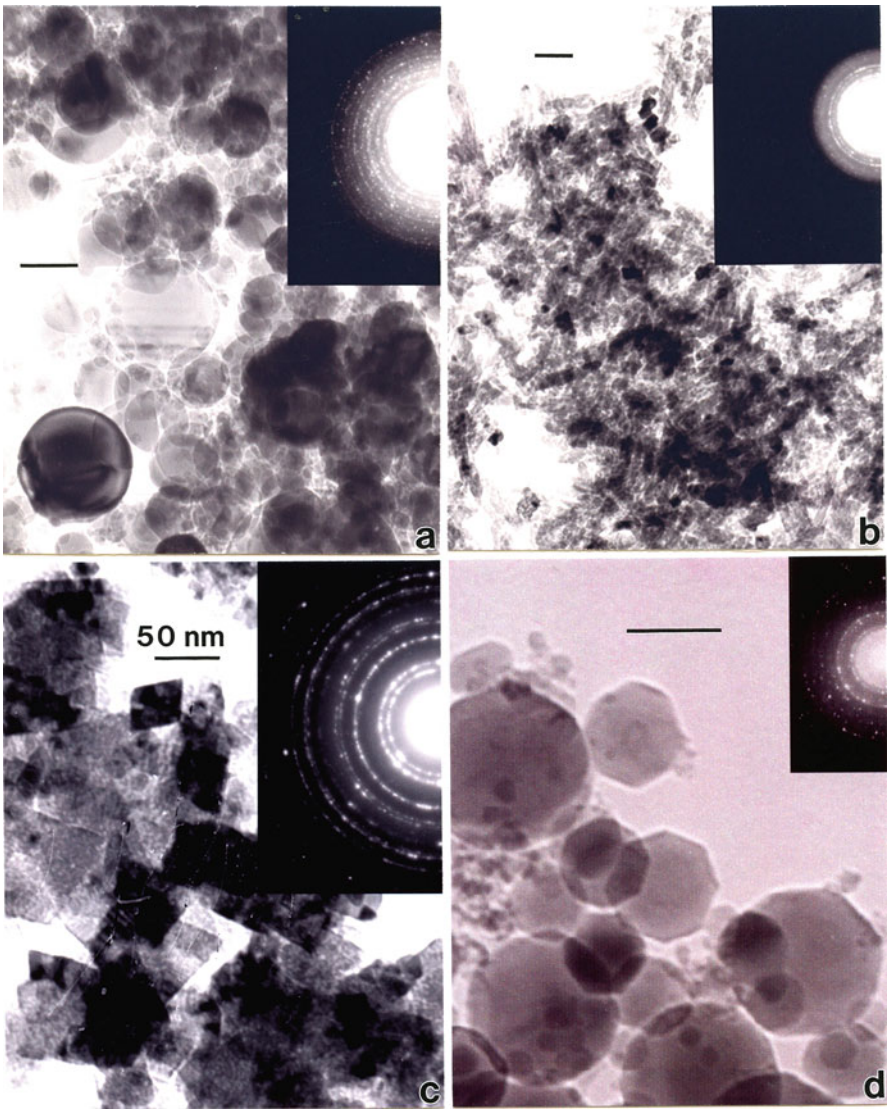


Fig. 3 Examples of nanoparticulate aggregates (or agglomerates) observed in the TEM. Corresponding SAED pattern half is shown in insert. (**a**) TiO_2 (rutile). (**b**) TiO_2 (rutile needles). (**c**) NiO. (**d**) Fe_2O_3 (hematite). Magnification marked as shown in (**c**)

oriented islands, many exhibiting specific crystal facets or edges. Figure 7c and d illustrate similar distributions of In nanoparticles vapor grown on NaCl (001) substrates. Figure 7c is a TEM image while Fig. 7d is a corresponding SEM image of the In single-crystal nanoparticles on the NaCl (001) surface. The insert

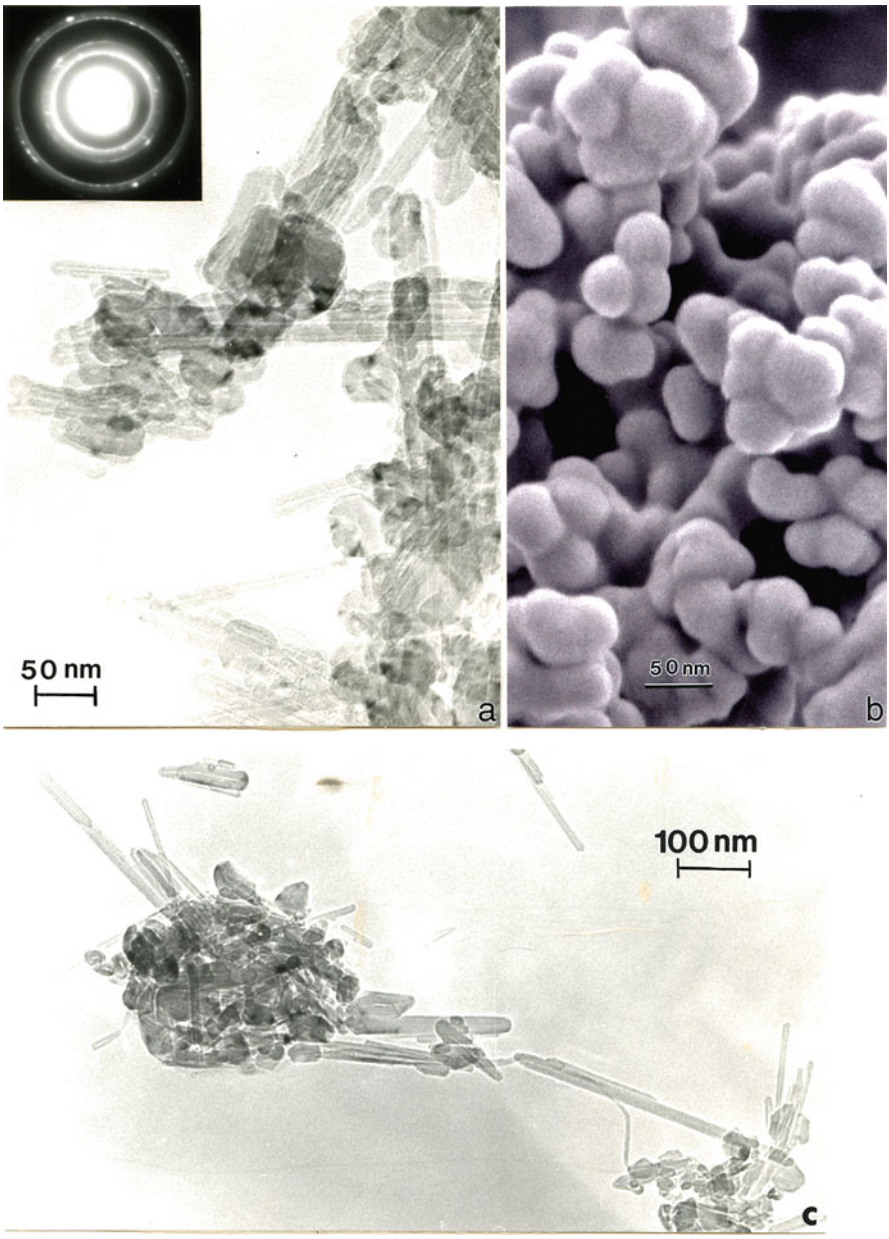


Fig. 4 Carbon nanoparticles. (**a**) TEM image of multiwall carbon nanotube and multi-concentric or multilayered fullerenes formed by propane gas combustion. (**b**) SEM image of wood soot aggregates. (**c**) TEM image showing propane combustion-produced multiwall carbon nanotube and fullerene aggregates. Insert SAED pattern in (**a**) shows graphite crystal diffraction rings and spots

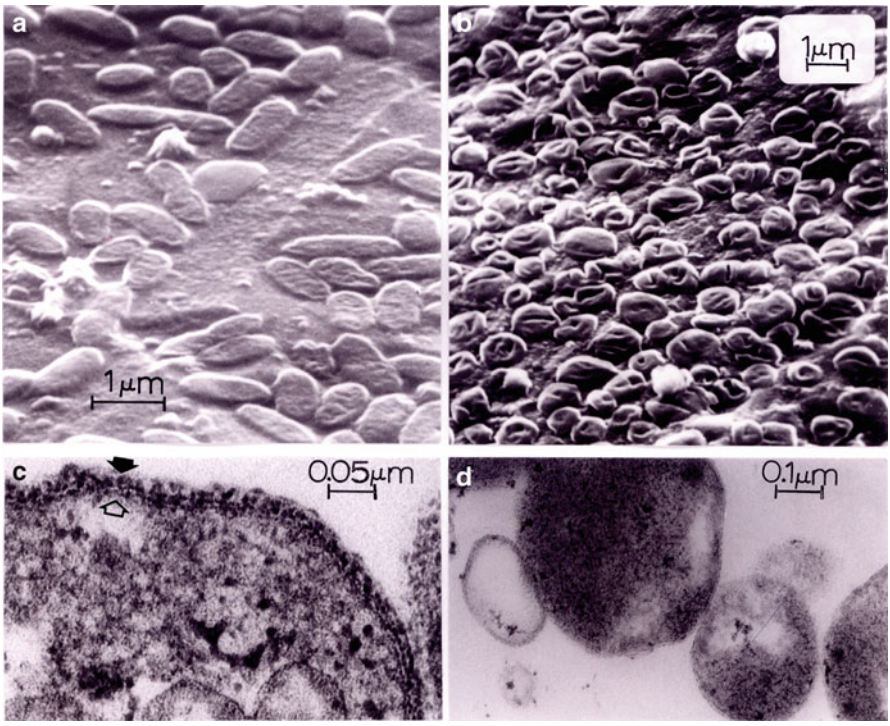


Fig. 5 SEM (**a**) and (**b**) and TEM (**c**) and (**d**) images of bacteria. (**a**) SEM image of *T. ferrooxidans* or mineral surface. (**b**) *Sulfolobus* on mineral surface (SEM). Note drying in air distorts shape. (**c**) TEM section image of bacteria in (**a**). Arrows denote cell wall structure. (**d**) TEM section image of bacteria in (**b**). (**c**) and (**d**) sections stained with phosphotungstic acid to produce mass-thickness contrast

in Fig. 7d shows a ball model representing the atomic nature and crystal faceting of these vapor-grown In nanoparticles.

Figure 8a and b illustrate two types of cubic or cuboidal precipitates in a solid matrix observed in the TEM. Figure 8a shows $Cr_{23}C_6$ cubic (10.6 Å lattice parameter) precipitates in a Co-Cr alloy matrix, while Fig. 8b shows closely spaced γ' (Ni_3Nb) cuboidal precipitates in a Ni-Cr alloy matrix. Because of the regularity and spacing of the precipitates in Fig. 8b, they have a more pronounced effect on the mechanical properties than the irregularly spaced carbide precipitates in Fig. 8a. In contrast, Fig. 8c shows regular arrangements of deformation twins or twin faults coincident with the {111} planes in fcc Ni following shock deformation. These have a thickness of roughly 10 nm and have a significant effect on the nickel matrix mechanical properties. Finally, Fig. 8d shows dynamically recrystallized nanograin structure in a tantalum shaped-charge component or jet fragment. While the TEM

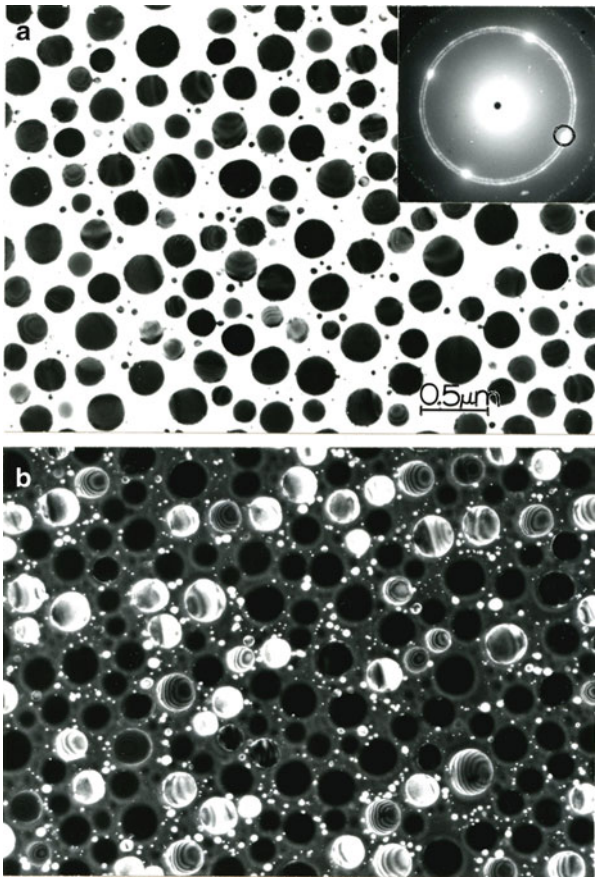


Fig. 6 Single-crystal nanotin particles growing from the vapor on NaCl (001) single-crystal surface. Extraction replica (carbon film supporting the Sn nanoparticles (Murr 1991)). (**a**) TEM bright-field image and SAED pattern insert. (**b**) TEM dark-field image of (**a**) using diffraction spot circled in SAED pattern insert in (**a**)

image in Fig. 8d is made in a thin film section, this represents a bulk, 3D, equiaxed nanocrystal or nanograin material, unlike the thin film nanograin material represented in Fig. 7a.

Figure 9 completes the TEM examples of nanomaterials structures and shows the variations in nanoprecipitates which are characteristic of heat treatment or aging in aluminum alloys as noted in the Introduction of this chapter. Figure 9a shows late-stage Guinier–Preston zones composed mainly of Cu atoms coincident with {100} planes which give rise to strain-field contrast along these coherent nanoinclusions. In contrast, Fig. 9b shows later-state precipitates of various compositions, sizes, and morphologies, although all are in the nanosize range (~20–60 nm). Dislocations interacting with these nanoprecipitates are also observed in Fig. 9b. Figure 9c shows fine Al_2Cu needle precipitates coincident with crystal planes in Al-6061 alloy and referred to as a Widmanstätten pattern. These needles or nanocrystal rods are roughly 4–20 nm in diameter and 200 nm in length. Other, larger and more irregular precipitates are also intermixed with the precipitate needles in Fig. 9c. All of the varieties

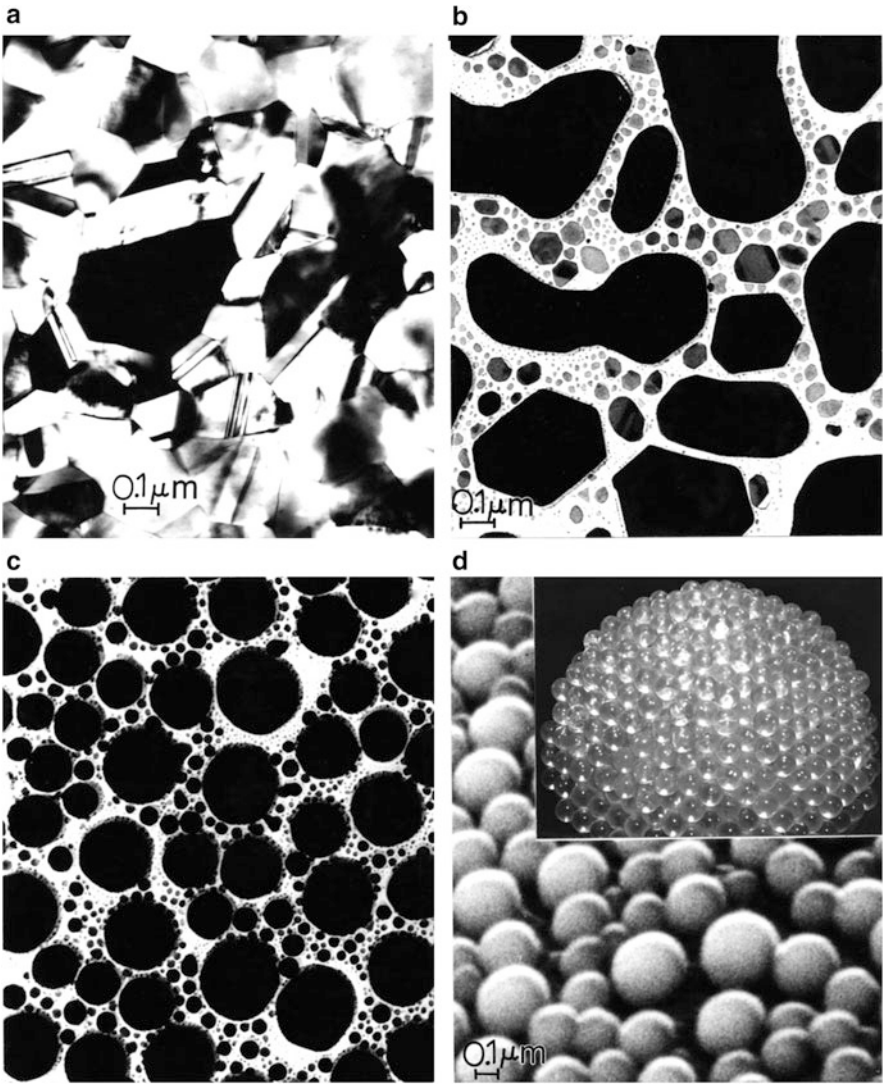


Fig. 7 Indium single-crystal nanoparticles vapor-deposited onto NaCl (001) single-crystal substrate. (**a**) Continuous thin film grown by Sn island nucleation and coalescence illustrated in (**b**). (**c**) Shows TEM image of nano-tin particle nucleation and growth at different conditions than (**b**). (**d**) SEM image of (**c**). Insert shows ball model of crystal structure (atom structure)

of nanoparticulates and nanomorphologies in Fig. 9 are nanophases in a solid-state fcc Al-6061 alloy matrix.

While SEM and TEM are among the most direct methods for characterizing most nanomaterials – their morphologies, nanostructures, chemical compositions, crystal structures, and crystallography – there are other methods for directly

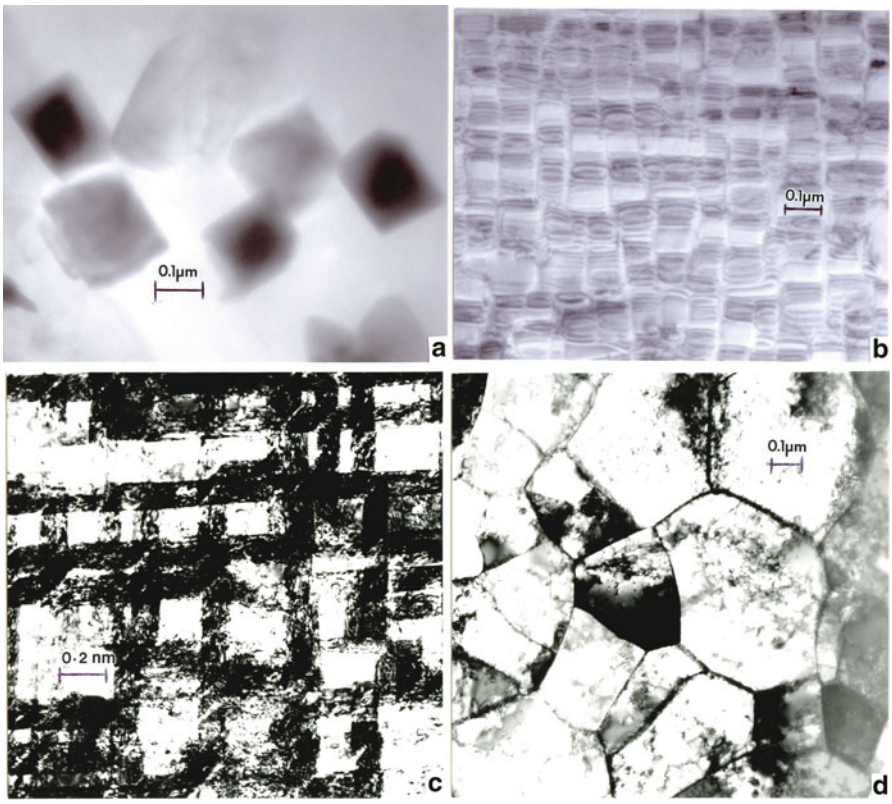


Fig. 8 Examples of nanoparticles/nanophases in bulk (or 3D structures) observed by TEM in thin film sections. (**a**) $Cr_{23}C_6$ cubic precipitates in Ca-Cr (fcc) matrix. (**b**) Ni_3Nb cuboidal precipitates in Ni-Cr (fcc) matrix. (**c**) Thin deformation twins coincident with {111} planes in shock-loaded Ni. Surface orientation is (100). (**d**) Nanograin structure in shaped-charge jet fragment of Ta (fcc)

observing nanostructures and even atomic structures. These include variations of nanoprobes, including the direct observation of atomic structure of nanoprobed surfaces. The oldest and most limited of these techniques involves *field-ion microscopy*, invented in the mid-1950s by Erwin Müller who also invented the *field-electron emission microscope* in the late 1930s. This technique, illustrated schematically in Fig. 10, produces projection images of the atomic structure of a whisker (single-crystal) surface as shown in Fig. 10a (Müller and Tsong 1969). It operates as a pin-hole camera by allowing helium ions to be polarized by an electric field established between the whisker end and a fluorescent screen coating on a thin, transparent, tin oxide semiconducting coating. The helium gas atoms "hop" over the whisker surface and lose an electron, becoming a positive ion and traveling along an electric field line corresponding to atoms "protruding" from, or characterizing the edge of, atomic planes as shown in Fig. 10b. These He ions (+) form a dot of light on striking the phosphor screen, forming an image corresponding

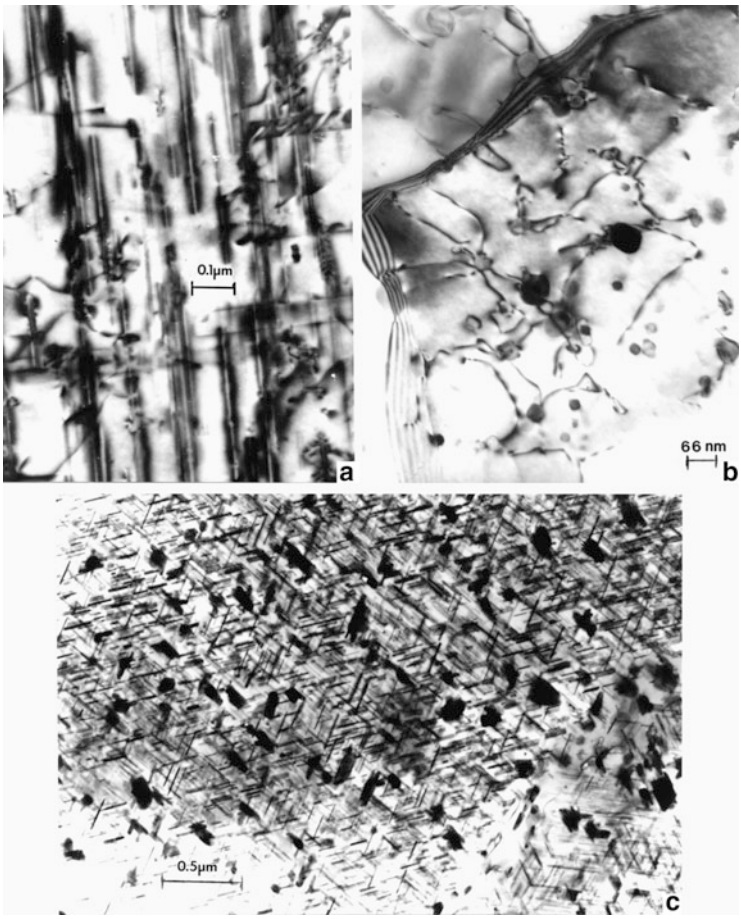


Fig. 9 Examples of nanoprecipitates/nanophases in Al-6061 alloy matrix. TEM observations. (**a**) Guinier–Preston zones coincident with {100} planes. (**b**) Globular precipitate with extended heat treatment relative to (**a**). (**c**) Widmanstätten structure composed of precipitate nano-needles and other globular precipitates

to an atom position. Because these surface atoms are "vibrating" as a consequence of intrinsic thermal vibrations, a coolant is required to form recognizable images. Correspondingly, the lower the coolant temperature, the more clear the image on the fluorescent screen, which is photographed below the screen in Fig. 10a.

Figure 11 shows a sequence illustrating the features of imaging a whisker tip or emitter end-form for a single-crystal whisker of fcc iridium. In Fig. 11a, the whisker, formed by selective etching by dipping into an acid (Murr 1991) and spot-welded onto a wire holder, is illustrated in the SEM image of Fig. 11b. The whisker or emitter axis is [001]. This orientation is shown in a ball model representing Ir atoms arranged in the [001] orientation in Fig. 11c, while the

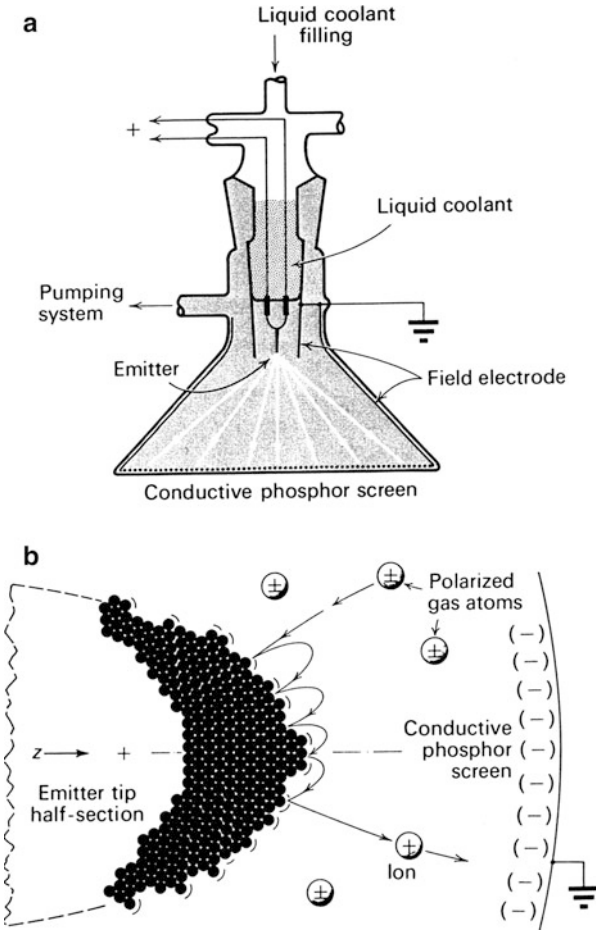


Fig. 10 Schematic view of the field-ion microscope (**a**) and FIM image formation (**b**). In (**a**) the FIM emitter (whisker specimen) is connected to a high-voltage (DC) position terminal, while the screen is (−) (or grounded). In (**b**) helium gas (imaging gas) is ionized (+) on the whisker surface at atom sites creating high electric field lines which project the ions to form an image "dot" on the phosphor screen (From Murr (1991))

corresponding field-ion microscope (FIM) image is shown in Fig. 11d, using a liquid nitrogen (78 K) coolant (Fig. 10a). It can be noted in Fig. 11d that the iridium unit cell (001) crystal face, containing four corner atoms and 1 face atom, is observed. This feature is shown enlarged in Fig. 12 by the arrow and the small "box" which characterizes the (001) face side of the fcc Ir unit cell. It can be noted that the corner (Ir) atom image to the left of the white arrow in Fig. 12 is egg-shaped, and this distortion occurs because of the high electric field condition which exists at this unit cell corner. The other corner atoms are similarly distorted. Note that the face-centered atom is not visible as prominently as the corner atoms because the electric field intensity is notably less than at the unit cell corners. Other atoms highlighted along the symmetry edges are more "round" because of their position on the end-form surface. The magnification of Fig. 12 can be determined by measuring the dimension of the side of the box representing

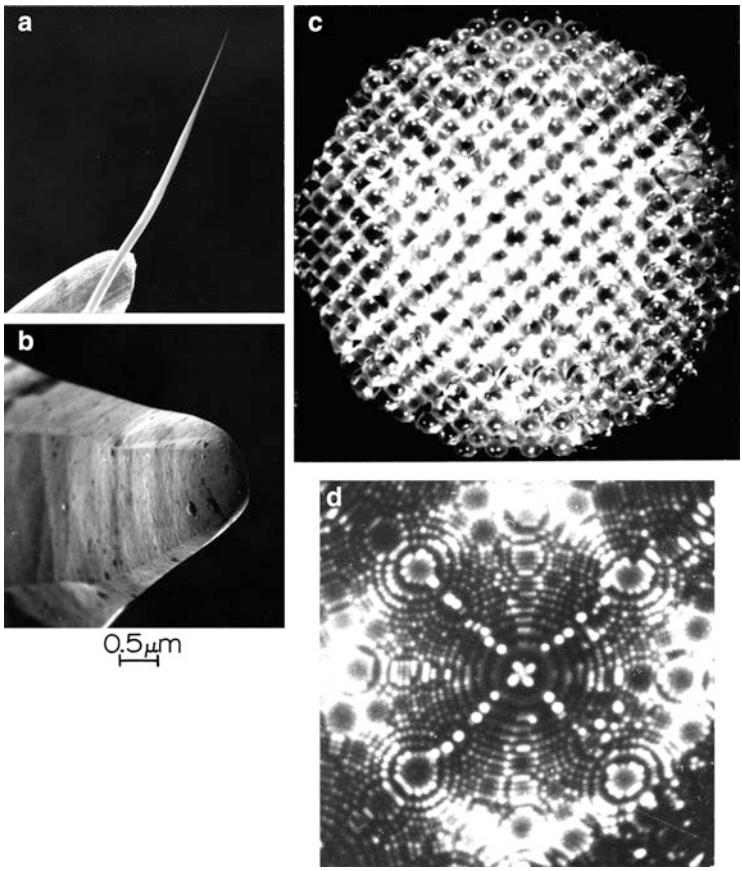


Fig. 11 FIM end-form (emitter) observed in the SEM ((**a**) and (**b**)). (**c**) shows a ball model of atomic structure viewed along [001]. (**d**) FIM [001] image for Ir (From Murr (1991))

the unit cell of Ir and dividing this real measurement by the Ir unit cell dimension ($a = 3.84$ Å; or 3.84×10^{-8} cm): ~10,000,000 times. The magnification (M) also depends on the tip or emitter whisker radius of curvature (R) and the distance from the screen (L): $M \sim L/R$.

Since the FIM images form by He-gas ions traveling along electric field lines emanating at atom sites, they are orthographic projections because electric field lines are always perpendicular to conducting surfaces. However, the stereographic projection is a good representation of this image projection as illustrated for [001] in Fig. 13a. Figure 13b illustrates the projection of this orientation along [001] shown by the arrow, with the faceted crystal at the center of the great sphere. Figure 14 shows an FIM image projection for this orientation for platinum (fcc) observed using a liquid helium coolant (2.3 K) rather than liquid nitrogen as in Fig. 12. Although the magnification of Fig. 14 in contrast to Fig. 12 is different, Fig. 14 is a clearer image as a consequence of reduced thermal vibration of the

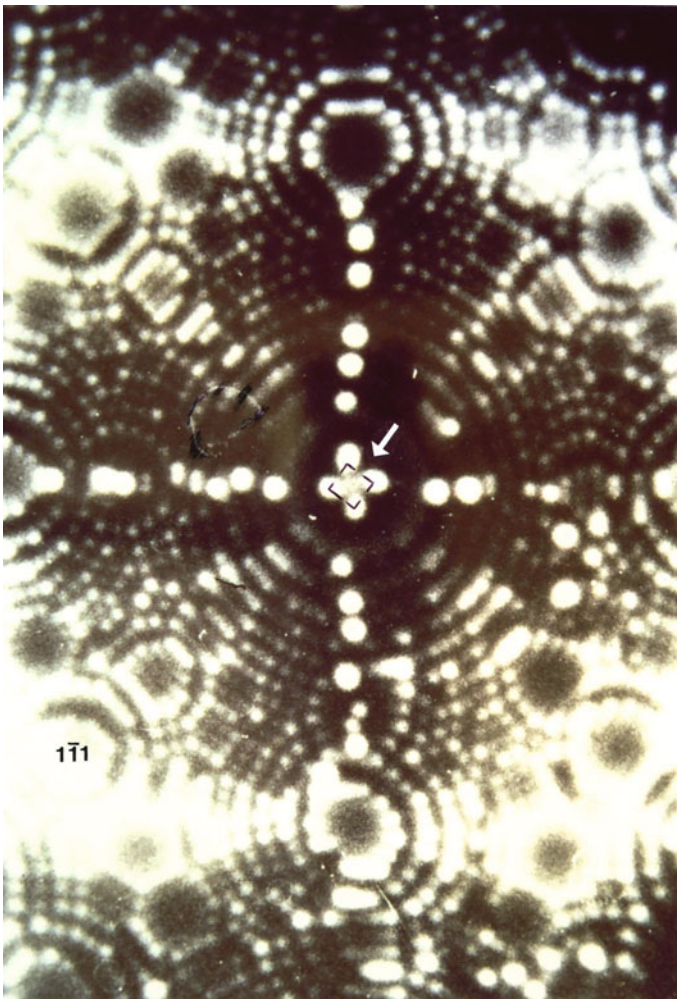


Fig. 12 Magnified view of [001] Ir image in Fig. 11 (**d**) showing Ir-fcc unit cell at *white arrow* and enclosed *box*

atoms. Figure 15 shows a similar [001] FIM image for Ir cooled with liquid helium, while Fig. 16 shows a [001] FIM image for tungsten cooled with liquid helium.

Although Figs. 12, 13, 14, 15, and 16 illustrate beautiful images of atomic structure, available for more than half a century, the FIM is applicable only to viewing refractory metal whisker surfaces, for the most part, and is thereby very limited in characterizing nanomaterials.

In contrast to FIM, the ability to create images of an atomic surface by using "whiskers" to probe the structure along x-y coordinates was developed first by applying an electrical (field) bias to allow the tip of the probe to ride above the

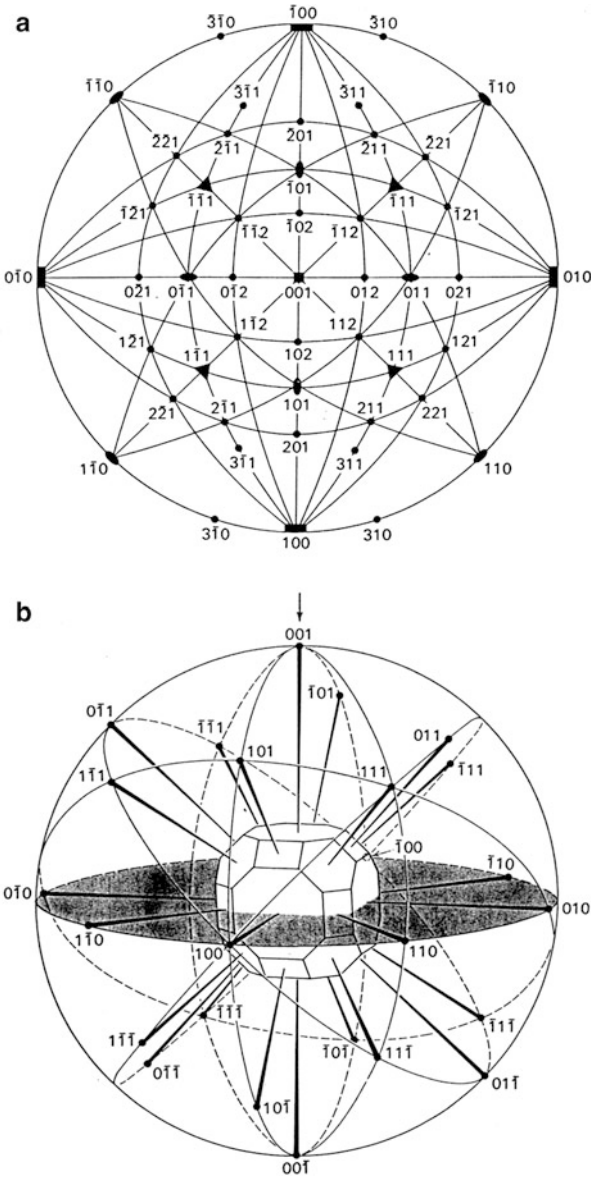


Fig. 13 (**a**) Cubic crystal stereographic projection onto [001]. (**b**) Stereographic reference sphere showing pole projections normal to crystal planes shown as fecetus in the crystal oriented at the sphere center. Arrow at top is a view of the [001] stereographic projection in (**a**) (From Murr (1991))

surface at a prescribed height, a technique referred to as scanning tunneling microscopy (STM). This was invented by Heinrich Rohrer and Gerd Binnig near the end of the 1970s. About 10 years later, Rohrer and Binnig along with C.F. Quate eliminated the tip bias and simply moved it as a direct surface probe or profilometer, thereby creating the atomic force microscope (AFM). Binnig and Rohrer received the 1986 Nobel Prize in physics for their work. Figure 17a compares these two

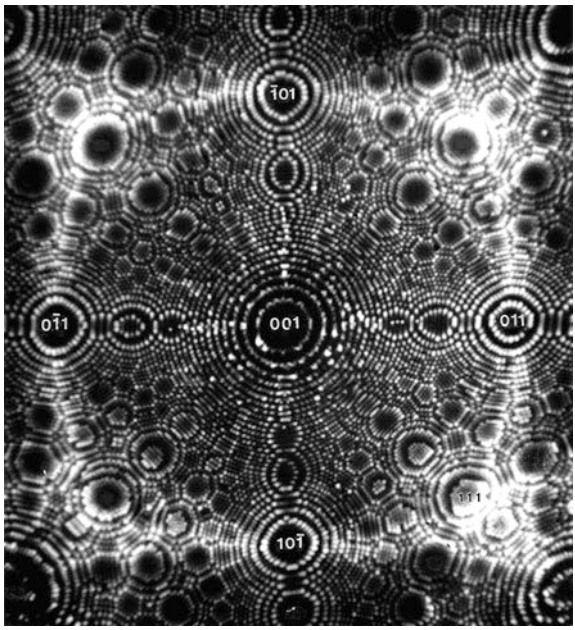


Fig. 14 FIM [001] image for platinum. Some plane normals or poles are indexed according to Fig. 13a (Courtesy of O. Nishikawa)

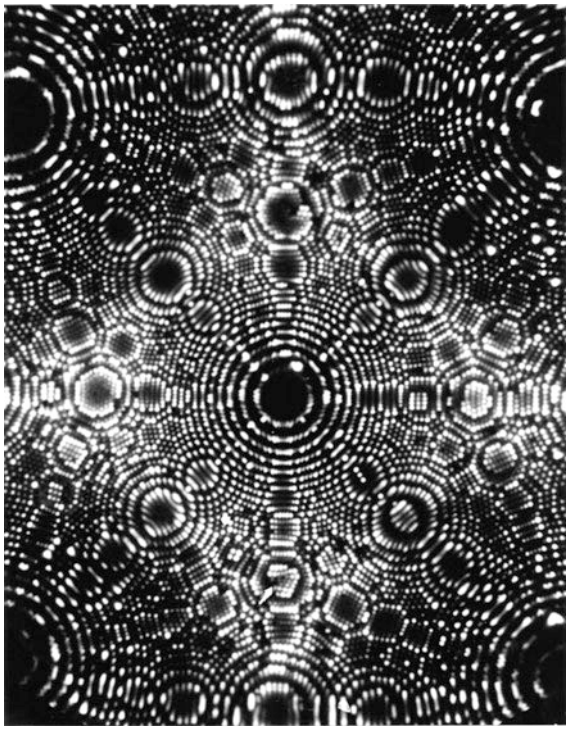


Fig. 15 FIM [001] image of iridium (Courtesy E. W. Müller (deceased))

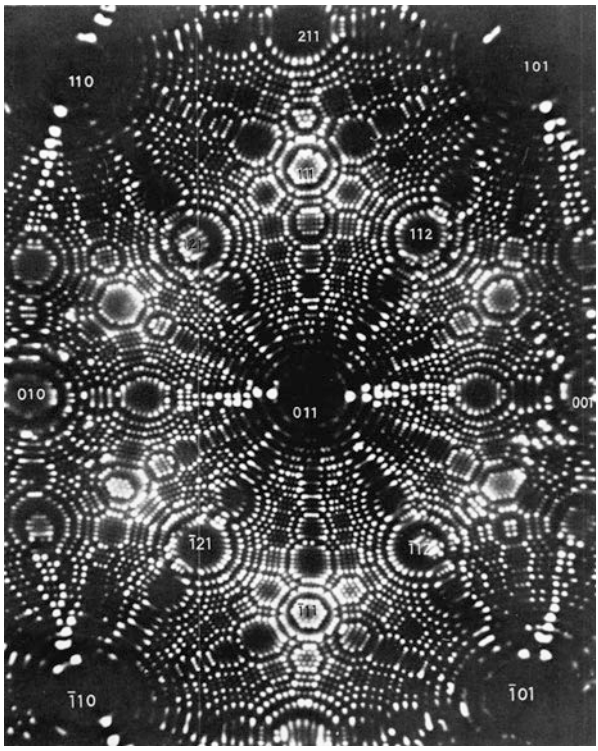


Fig. 16 FIM [001] image for tungsten (Courtesy of E. W. Müller (deceased))

characterization techniques in simple schematic form, while Fig. 17b illustrates an example of the atomic surface structure of a highly oriented pyrolytic graphite lattice imaged at room temperature in a commercially available AFM. Both STM and AFM instruments have been available commercially, and improving simplicity in achieving surface resolution for a wide range of materials is contributing to their ability to make increasingly meaningful images (Magonov and Whangbo 1996; Chen 2007; Bowen and Hilal 2009).

While this chapter has provided a brief overview of nanostructure and nanomaterials characterization using SEM, TEM, FIM, STM, and AFM, it has, more importantly, illustrated various types and classifications of nanomaterials. In particular, Figs. 2, 3, 4, 5, 6, and 7 illustrate a variety of effectively "free" nanoparticulate materials, even biological materials (Fig. 5) which represent primary nanoparticles having a fairly uniform nanodimension or diameter, nanorods or nanotubes having either a nano-diameter and nano-length, or a length which transcends the nano-regime (>1 μm). In contrast, Figs. 8 and 9 illustrate primary nanoparticles or nanophases in a solid matrix or crystal matrix or a 3D solid composed of nanograins or nanocrystals. These examples will form the basis for describing nanomaterials in relation to nanodimensions: 1D, 2D, 3D.

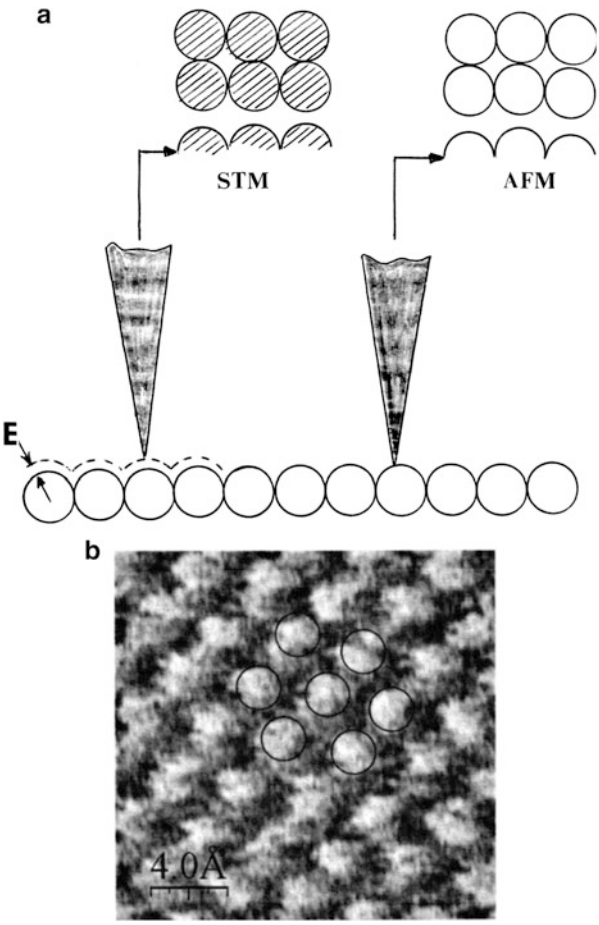


Fig. 17 Schematic views of STM and AFM (**a**). AFM image showing highly oriented pyrolytic graphite (carbon) atoms (**b**) (Courtesy of Nanotec Electronica)

Nanostructure Classification Issues: Dimensionality, Contiguity, and Functionality

Nanoparticles, their structure and geometry as it relates to polyhedral formation, were treated in some detail in chapter "► Chemical Forces: Nanoparticles" of Part III. This included clusters of various types and the creation of unit cells with clusters acting as superatoms and occupying lattice sites. In this sense, perhaps these clusters might represent zero dimension, but in reality they have dimension, especially nanodimension.

In classifying nanomaterials or nanostructures, dimensionality is often a defining issue: nanostructures being those with at least one dimension <100 nm. But nanostructure materials are often considered even with dimensions ranging as high as 700–800 nm since such structures are able to cause diffraction or

interference effects within the optical spectrum (at the red portion). Correspondingly, when a dimension becomes too small, the nanostructure begins to assume special quantum effects and exhibits quantum properties, especially electrical or optical (or nano optoelectronic properties). Even simple considerations such as quantum wells and related quantum confinements which influence the application of the Schrödinger equation or the deBroglie momentum–wavelength relationship implicit in the brief overview provided in chapter "► A Brief Introduction to Quantum Mechanics" of Part III invoke dimensional changes to achieve sizes <6 nm, such as the creation of quantum dots, which modify electron energy levels causing a general shift in the band gap, especially for semiconductor quantum dots. Certainly other dimensional regimes such as grapheme sheets ~0.3 nm thick experience similar phenomena and invoke related quantum properties.

Ideally, there are two functional classifications of nanostructured materials: free and contiguous, representing single or multiphase regimes. Free phase regimes are represented by nanoparticles and to some extent nanoclusters and aggregates. These can be unisize, 3D nanoparticles with a consistent dimension; thin sheets or nanoplates (or discs) having one nanodimension; or nanorods, nanotubes, or nanofilaments which can have all three nanodimensions, or only one, such as a nanorod or nanotube diameter, with a length >> 1 μm. Similarly, nanofilms, either free (such as graphene) or grown on a substrate, can have only a thickness dimension as a nanodimension as a single-crystal and two dimensions if the thin film is polycrystalline with a grain size in the nanodimensional regime. Such films can even be considered contiguous 3D nanocrystals in contrast to a bulk solid composed of nanocrystal grains having nanodimensions. In this context, nanoparticles in the form of dispersed phases with a uniform diameter, or as rods, tubes, or plates having nanodimensions, as well as precipitates or transformed phase regimes or other nano-regimes in a solid matrix, especially a crystalline matrix, become contiguous nanostructures. Many of these classification features are implicit in the nanostructures illustrated in Figs. 2, 3, 4, 5, 6, 7 and 7a, 8, and 9. These nanostructure features are illustrated ideally and schematically in Figs. 18, 19, and 20. Figure 18 shows cube and spherical nanoparticles with dimension, d in the upper left with rods and tubes to the right, having dimensions d and 1, where one or both can be a nanodimension; or t_x and t_y nanodimensions while 1 can be either nano- or micro-dimensional. Discs and plates with nanodimension t are shown, and these can also have sheet-plane dimensions either in the nano-range or larger. The bottom schematic shows quantum rods (wires) or dots on a substrate, and this configuration can also be assumed to accommodate the thin film or sheet above to represent a thin nanofilm deposited on a substrate. In addition, the sheet (with thickness t) in Fig. 18 can also form a contiguous solid such as the composite steel laminates characterizing the Damascus and Japanese swords discussed in chapter "► Serendipitous Nanotechnology in Antiquity," as well as other nanolaminate structures.

Figure 19 shows 3D nanoparticles, rods, or thin phases in a solid crystal section, a contiguous, 3D solid. In contrast, Fig. 20 shows two configurations of contiguous

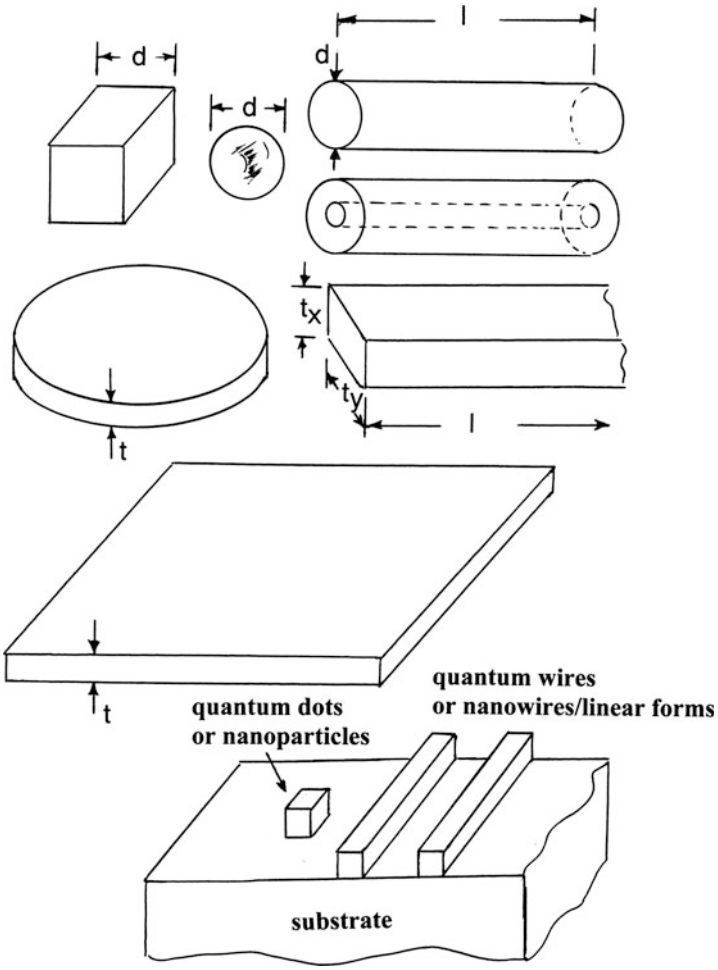


Fig. 18 Simple nanoforms (cubes, spheroids, rods, tubes, discs, plates, etc.) or nanostructures representing free "particles" and semi-contiguous structures such as nanoparticles or quantum dots or wires on a substrate (*bottom*). Larger nanostructures can create interference or diffraction phenomena as discussed in chapter "▶ Electromagnetic Color and Color in Materials"

or continuous nanophases composing a solid section, such as a nanocrystalline solid composed of nanograins. The upper left schematic in Fig. 20 illustrates either a microcrystalline solid or a nanocrystalline solid at the upper nano-range (<1 μm) where the dimensions of the grain boundary or phase boundary region (Δt) are small in contrast to the grain (matrix) dimension (D): $\Delta t << D$. The lower-right schematic (Fig. 20) depicts a 3D, nanocrystalline solid where the grain boundary thickness, Δt, becomes commensurate with the grain diameter, D: $\Delta t \leq D$. These two circumstances pose considerable differences in behavior

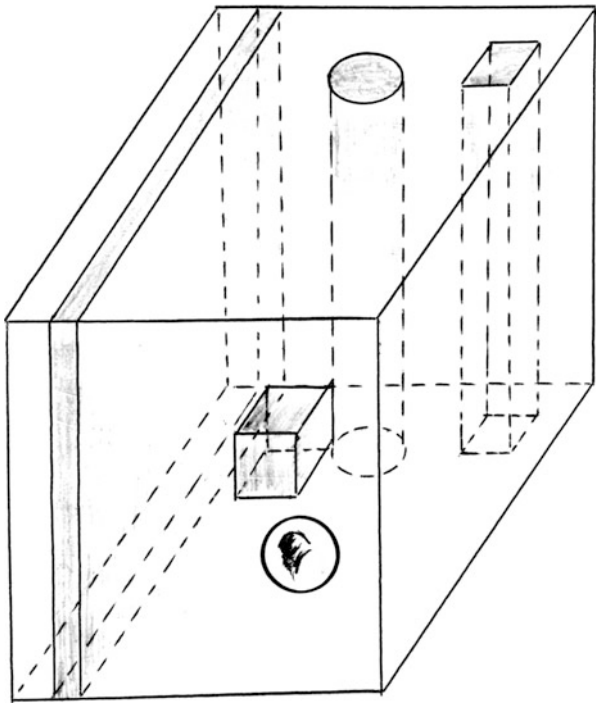


Fig. 19 Nanoparticles or nanoforms in a contiguous, solid crystal

(or properties), especially mechanical properties. This feature is also implicit in considering amorphous solids which might be considered the end point of a nanomaterial where there is little or no distinction between contiguous nanocrystals or grains and their separating regime (or grain boundaries). Murr (1980) considered this situation in the context of how the Hall–Petch relationship for polycrystalline metals

$$\sigma = \sigma_o + K/\sqrt{D} \tag{1}$$

might contribute to, or represent, the hardness or yield stress of an amorphous (or glassy) metal, which are usually significantly greater than a polycrystalline, engineering metal. Note in Eq. 1 that σ_o and k are constants.

Figure 21 provides a perspective for this situation and a 2D correlation with the nanocrystalline context shown in Fig. 20. Figure 21a illustrates a crystal section of a ternary alloy (three atom species), while Fig. 20b shows this same section disordered, or having an amorphous (noncrystalline) structure. However, Fig. 21c shows the disordered structure in Fig. 21b to have small, ordered units, but not crystal units or unit cells. Correspondingly, Fig. 21d shows a polycrystalline structure which represents a 2D rendering of the upper-left schematic of Fig. 20. In contrast, Figs. 21e and f are 2D renderings characteristic of the lower-right schematic of

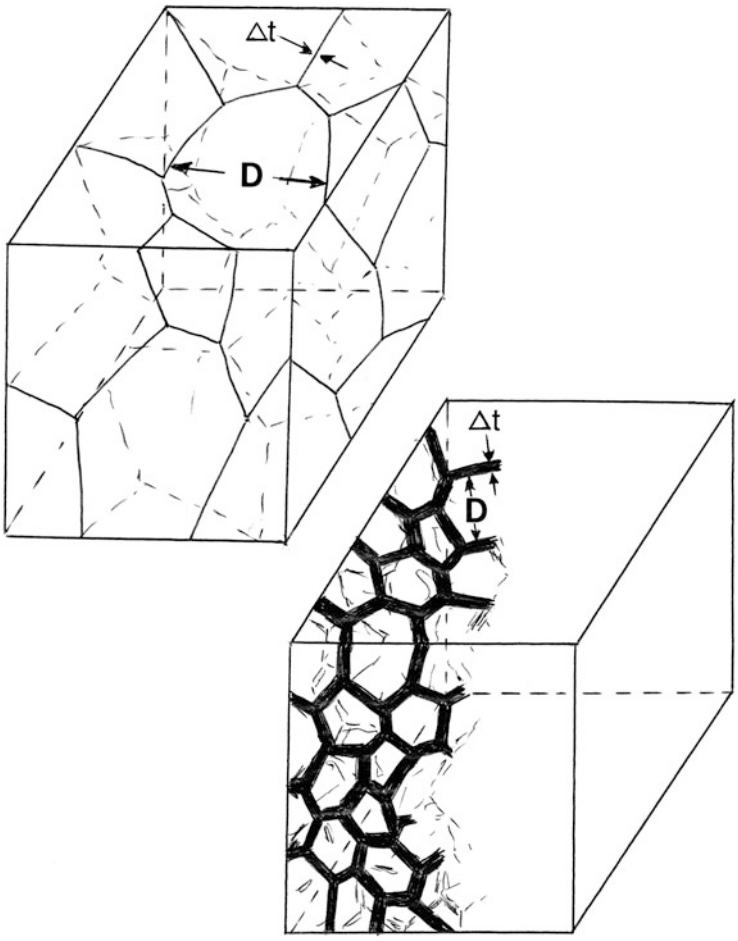


Fig. 20 Contiguous nanocrystalline structures or nanograin structures: coarse grain *upper left*; nanograin *lower right*

Fig. 20. Finally, Fig. 21g shows a 2D schematic for nanograin structure commensurate with Fig. 21e and the lower-right schematic in Fig. 20.

In Fig. 21g the nanograins are depicted as solid circles while the grain boundary region is indicated by open circles. It can be noted that the grain boundary region is represented ideally as an amorphous-like region similar to Fig. 21c, ignoring the three different atomic species. One of the related features implicit in Fig. 21g is that the entropic features of the nanograins (dark) are certainly different from the disordered grain boundary or inter-grain regime. Correspondingly, the energy would be intuitively higher in this regime in contrast to the ordered (crystalline) nanograin regime (dark circles).

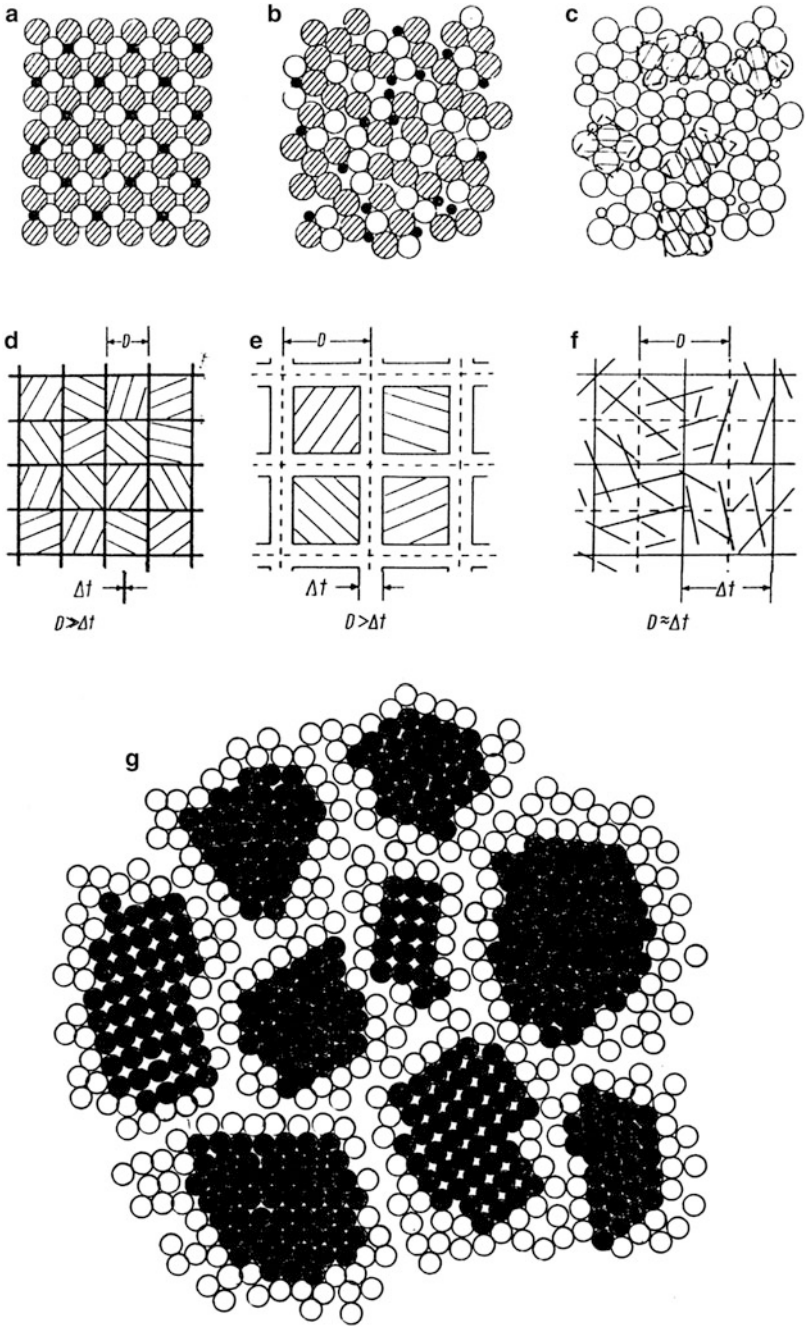


Fig. 21 Connections between crystal structure, amorphous structure, and nano(crystal) structure. (**a**) Crystal structure for ternary alloy. (**b**) Disorder (amorphous) structure. (**c**) Disorder–order

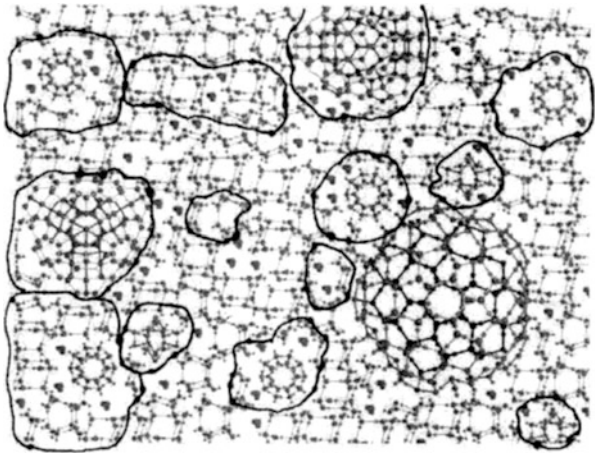


Fig. 22 Proposed structure of liquid water from Roy et al. (2009). Nanounits or nanoclusters are outlined in black

It can be observed in retrospect that the schematic representations shown in Figs. 18, 19, 20, and 21 depict many nanostructures in the context of classifications. However, it may be useful to consider a few other nanostructures. One would be colloidal suspensions: nanoparticles suspended in a liquid. The other would be nanostructures associated with a liquid. Just as a glass or glassy solid depicted in Fig. 21c as an amorphous or disordered solid might be viewed in the context of regular units of atoms dispersed in this disordered matrix, so also a liquid might be viewed in this context. One of the more relevant examples is water, or more appropriately the structure of water: H_2O or H : O : H. In contrast to the so-called nanostructure units depicted in Fig. 21c, Fig. 22 shows nanostructure units in water as conceptualized by Roy et al. (2009). In three dimensions, these units can be either regular or irregular polyhedra, and they will change the structure and dimension with temperature. The practical consequences of this, as implicit in the work of Roy et al. (2009), are that small concentrations of medications, for example, may interact differently with water at different temperatures, i.e., "some relevance to homeopathy." Indeed, this is perhaps implicit in the so-called Chinese medicine anecdotes passed down for more than 1,000 years, one of which dictates that for medications to be effective requires them to be taken with "warm" water, in contrast to "cold" water, where even intuitively from Fig. 22 it might be expected that the nanostructures or nanostructure units would be different. The reader might look at a recent book by Gerald Pollack which discusses the "fourth phase of water" – a so-called exclusion zone that forms next to many submersed materials, which has a lot of charge and differs from bulk water in character (Pollack 2013).

Fig. 21 (continued) structure (shaded units). (**d**) Regular polycrystalline material structure. (**e**) Nanograin structure. (**f**) Amorphous structure. (**g**) Nanograin structure consisting of nanocrystals (*dark circles*) separated by an interfacial phase or grain boundary region (*open circles*) (Adapted in part from Murr (1980))

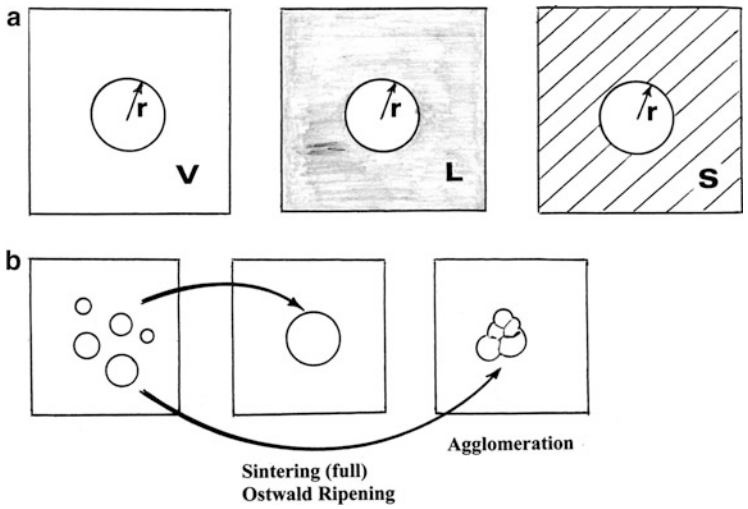


Fig. 23 (**a**) Solid (spherical) particle nucleated and growing in vapor (*V*), liquid (*L*), or solid (*S*) phase. (**b**) Energy reduction by particle sintering, Ostwald ripening, or agglomeration

Similarly, in the work of Roy et al. (2009), they make the most provocative statement which in fact transcends water: "Structures change properties vastly more easily and dramatically than chemistry changes them." This is truly an important consequence of essentially all materials structures and is especially pertinent in considering nanomaterials and nanostructures.

Certainly, this brief treatment does not encumber all nanostructures or classifications, but it provides an overview. There is a reasonable correspondence between Figs. 19, 20, and 21 and the observations of nanostructures illustrated in Figs. 2, 3, 4, 5, 6, 7, 8, and 9. Many other examples have been provided in chapter "▶ Chemical Forces: Nanoparticles" and related parts and chapters in this handbook.

Interfacial Phenomena in Nanomaterials

In discussing cuboctahedral metal nanoparticles in chapter "▶ Chemical Forces: Nanoparticles," it was noted in Table 1 that when the cluster or particle diameter is very small and contains only one shell of atoms around a central atom, the percent of surface atoms is >90 %. Correspondingly, the surface free energy dominates since in effect the only real particle volume is a single (central) atom. Even for nucleating or growing particles in some phase, vapor (V), liquid (L), or solid (S) as illustrated in Fig. 23a, the total (Gibbs free energy) energy is ideally given by

$$\Delta G = \frac{4}{3}\pi r^3 \, \Delta G' + 4\pi r^2 \gamma_I \tag{2}$$

where r is the particle radius, $\Delta G'$ is the volume free energy, and γ_I is the interfacial free energy. It can be observed in Eq. 2 that for very small nanoparticles the interfacial free energy term far exceeds the volume free energy term.

Surface energy or surface (or interfacial) free energy is the energy required to create new surface:

$$\gamma_I = \left(\frac{\partial G}{\partial A}\right) n_i, T, \quad P \tag{3}$$

where G characterizes the Gibbs free energy term, A is the interfacial area, n_i the is number of surface atoms (see, e.g., ▶ Table 1, chapter "Chemical Forces: Nanoparticles"), T is the temperature, and P is the attendant pressure. As illustrated in Fig. 23b, a regime of particles can reduce their energy by three basic processes: sintering, where particles are caused to contact and ideally form a single particle having the same volume as the combined particles (volume conservation); *Ostwald ripening* where in a non-contact process the smaller particles can "dissolve" (or vaporize) and "condense" by diffusion on a growing (larger) particle; or simply agglomerate, forming irregular arrays. Examples of sintering and agglomeration of nanoparticles are illustrated in Figs. 2 and 4b, among others. Other agglomerations can occur as shown in Figs. 4a and c, but these involve electrostatic or van der Waals attraction and there is no physisorption or intrinsic bonding of matter. Such agglomerates do not measurably reduce the associated surface free energy.

In the case of Ostwald ripening, for example, particle curvature is involved since small particles involve small curvatures (for ideal spheres). In contrast to a "flat" or infinite (∞) surface, the difference in chemical potential can be expressed by

$$\mu_p - \mu_\infty = \Delta\mu = \Delta\mu = \gamma_I \Omega \left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{4}$$

where μ_p is the chemical potential for the particle having radii of curvature R_1 and R_2 and Ω is the atomic volume. For an ideal, spherical particle, $R_1 = R_2 = r$ and Eq. 4 become

$$\Delta\mu = 2\gamma_I \Omega / r, \tag{5}$$

often referred to as the La Place equation. Considering the equilibrium vapor pressure of spherical particles (Fig. 23a), the difference in chemical potential can be written as

$$\Delta\mu = kT \ln\left(\rho_p / \rho_\infty\right) \tag{6}$$

where k is the Boltzmann constant, T is temperature, and p_p and p_∞ are the particle and flat surface equilibrium vapor pressures. Similarly, in a liquid or solid phase (Fig. 23a), the solubility becomes

$$\Delta\mu = kT \ln \left(S_p/S_\infty\right) \tag{7}$$

where S_p and S_∞ represent the corresponding solubilities. For small spherical particles,

$$\ln \left(pp/p_\infty\right) = \ln \left(S_p/S_\infty\right) = 2\gamma_I\Omega/kT \tag{8}$$

It is observed that the vapor pressure and solubility of small (spherical) particles is higher than that for bulk material. Consequently, for Ostwald ripening, there is a net diffusion from small-to-larger particles where in effect the small get smaller and the large get larger. It is observed in Eq. 8 that this depends on the interfacial free energy and the "driving force" to reduce the overall energy contribution by the surface free energy or, in the case of nanocrystalline solids as illustrated in Fig. 20, the growth of nanograins. Even in contiguous solids such as the Widmanstätten structure in an Al-6061 alloy as shown in Fig. 9c, there is a certain nonequilibrium feature because of the large surface or interfacial free energy or interfacial energy component ($\sum\gamma_I\, \varsigma_I$) contributed by the densely packed nanoprecipitate needles.

It is apparent on examining Fig. 20 that in contrast to a single-crystal reference cube corresponding to those polycrystal cubes shown in Fig. 20, contiguous polycrystals add interfacial energy to the solid or volume energy and that as the polycrystals or grains become very small nanograins, the solid section becomes dominated by interfacial energy; that is, the interfacial area (A_I) within the volume element shown in Fig. 20 increases dramatically. For truly nanomaterials, the driving force to reduce this internal energy contribution becomes very large, and the nanograins will grow. In the case of nanocrystalline metals, this can become a very important issue because if processing to produce a nanomaterial is performance driven and performance declines because the nanograin structure grows significantly, then strategies to retain the nanograin structure or in fact any other nanophase structure become a crucial issue.

This is not unlike similar grain growth issues in other polycrystalline metal or alloy applications which have required materials engineering strategies to be developed in the past. A particular case in point is the grain growth in tungsten and tungsten alloy filaments which occurred decades ago. The small grains in the precursor filament wires grew during high-temperature filament operation until the grain boundaries became perpendicular to the wire axis. This formed a classical *bamboo structure* of aligned, cylindrical, single-crystal grains, and stresses in turning the filament on and off caused sliding on these normal grain boundaries, reducing the wire cross-section and creating high-resistance zones which caused localized failure at turn on. This grain growth was eliminated by adding pinning elements (such as oxides) which segregated to the grain boundaries and prevented them from growing. In addition, such segregation reduced the interfacial (grain

boundary) free energy and thereby reduced the driving force for grain growth in addition to the pinning action of the segregating component (Murr 1975). Julia Weertman (2012) has recently addressed similar issues in nanocrystalline metals and alloys, especially the retention of high-temperature properties of nanomaterials. In some respects, similar issues exist in stabilizing the metastability inherent in amorphous metals and alloys which crystallize at the so-called glass transition temperature, roughly 450 °C in many iron-based amorphous or glassy alloys such as 80 Fe-20 B. In the case of nanocrystalline or nanograin materials shown schematically in Fig. 20, grain growth can occur even at relatively low temperatures, and the application of stress also hastens grain coarsening (Weertman 2012).

References

Bowen WR, Hilal N (eds) (2009) Atomic force microscopy in process engineering. Elsevier, Oxford

Chen CJ (2007) Introduction to scanning tunneling microscopy, 2nd edn. Oxford University Press, Oxford

Gleiter H (2000) Nanostructured materials: basic concepts and microstructure. Acta Mater 48:1–29

Magonov SN, Whangbo M-H (1996) Surface analysis with STM and AFM. VCH Publishers, New York

Merica PD, Waltenberg RG, Scott W (1919) Heat treatment of duralumin. Bull Am Inst Min Metall Eng 150:913–949

Müller EW, Tsong TT (1969) Field ion microscopy: principles and applications. Elsevier, New York

Murr LE (1975) Interfacial phenomena in metals and alloys. Addison-Wesley, Reading

Murr LE (1980) A form of hall–petch law for the strength of glassy metals. Phys Stat Sol (a) 6l: K101–K107

Murr LE (1991) Electron and ion microscopy and microanalysis: principles and applications, 2nd edn. Marcel Dekker, New York

Pollack GH (2013) The fourth phase of water: beyond solid, liquid, and vapor. Ebner & Sons, Seattle

Roy R, Tiller WA, Bell I, Hoover MR (2009) The structure of liquid water; novel insights from materials research; potential relevance to homeopathy. Mater Res Innovat 9(4):577–608

Siegel RW, Hu E, Roco MC (eds) (1999) Nanostructure science and technology, Final report by WTEC panel, Sept 1999. Loyola College, Maryland. http://itri.loyola.edu/nano/final/

Siegel RW (1993) Synthesis and processing of nanostructured materials. In: Nastasi M, Nastasi M, Parkin DM, Gleiter H (eds) Mechanical properties and deformation behavior of materials having ultrafine microstructures, NATO-ASI, vol E233. Kluwer Academic, Dordrech, pp 508–536

Weertman JR (2012) Retaining nano in nanocrystalline alloys. Science 337:921–922

Wilm A (1911) Physikalisch-metallurgische untersachungen über magnesium haltige aluminum legierungen. Metallurgie: Zeitschritt fip der gesamte Hüttenkunde 861:225–230

Synthesis and Processing of Nanomaterials

Contents

Abstract

Nanoparticles and nanoparticle aggregates are produced as colloids and related nanocrystals by sol–gel and wet chemical synthesis as well as a variety of vapor, spray, and plasma processes, including flames. Vapor-phase material can be condensed or deposited on variously heated or cooled substrates, even crystalline substrates where films, islands, and quantum dots can be grown. The role of catalysts in promoting reactions is described. Sputter deposition and other physical vapor deposition (PVD) processes are discussed, including arc evaporation. Chemical vapor deposition (CVD), including reaction product production and collection, is presented. Molecular beam epitaxy (MBE) and atomic layer deposition are described. Collected nanopowders can be statically or dynamically consolidated to form bulk billets. Mechanical alloying and mechanochemical synthesis of nanomaterials are discussed along with electrodeposition, friction-stir processing (FSP), and equal-channel angular processing (ECAP) or extrusion as these apply to severe plastic deformation (SPD) processes to produce nanocrystalline bulk solids.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_46

Introduction

Because there are so many types or classifications of nanomaterials and structures, there are correspondingly a great variety of synthesis and processing routes to achieve them, especially in commercial quantities, some of which, such as black carbons, require tens of millions of tons globally on an annual basis. Thermodynamic modifications of pyrolysis processes (organic decomposition in the absence of oxygen) for the synthesis of black carbons are also used to produce varieties of graphites as well as carbon nanotubes. The production of nanometals such as gold and silver as described in Part XI, chapter "▶ 3D Printing: Printed Electronics," characterizes chemical reaction routes involving stabilization and reducing agents (Zhao et al. 2013). Similar chemical processes involve the sol–gel process, a wet chemical technique for producing colloidal particles (<30 nm) or particle networks (Brinker and Scherer 1990). Nanoparticles are also produced by a variety of physical processes involving spraying or deposition techniques, while bulk nanocrystalline materials generally require mechanical or deformation processing (Tjong and Chen 2004).

The so-called liquid synthesis routes such as *sol–gel processing* or *wet chemical synthesis* are especially conducive to producing metal oxide or ceramic nanoparticles such as those illustrated in ▶ Fig. 3 in chapter "Classifications and Structures of Nanomaterials." The sol–gel process offers a degree of control over composition and structure, even at the molecular level, through the production of a colloidal suspension (or sol) which is converted to a viscous gel or solid nanoparticulate. In this process, reactive metal precursors are initially hydrolyzed followed by condensation and polymerization reactions. Two mononuclear complexes of a metal, M, each comprising only one metal, M, can react to form a polynuclear complex consisting of two metal atoms. Condensation occurs via the so-called olation and oxolation reactions as follows:

$$\begin{aligned}&\text{OLATION}: \mathrm{M-OH + M-OH_2 \rightarrow M-OH-M + H_2O}\\&\text{OXALATION}: \quad \mathrm{M-OH + H-OM \rightarrow M-O-M + H_2O}\end{aligned}$$

or

$$\mathrm{M-OH + ROM \rightarrow M-O-M + ROH},$$

where R is an alkyl group, C_nH_{2n+1}. Under acidic conditions, 3D solid-phase networks form consisting of extended linear M – O – M chain molecules.

Wet chemical processes, including solid nanoparticle precipitation from supersaturated solution, homogeneous liquid-phase chemical reduction, or ultrasonic decomposition of chemical precursors, are particularly attractive because of their simplicity, versatility, and low cost. For example, titania (TiO_2) nanoparticle

production as illustrated in ▶ Fig. 3b of chapter "Classifications and Structures of Nanomaterials" by hydrolysis of $TiCl_4$ at low temperatures is expressed as

$$Ti\ Cl_4 + 2\ H_2O \rightarrow \mathbf{TiO_2} + 4\ HCl.$$

It is probably notable to indicate that nanoparticle synthesis of gold, silver, and copper in particular has been observed and studied in a number of biomimetic or "green" synthesis regimes including bacteria (Anastas and Warner 2000), fungi (Ahmad et al. 2003), and plants, where plant components uptake metal ions from the soil and produce small, colloidal, or neutral nanogold, copper, or silver crystals or polycrystal forms (MarChiol 2012; Das and Marsili 2010). In these processes the reduced, neutral metal nanoparticles can be recovered by simple combustion of the organic matrix. These regimes often produce very small crystalline, nanometal particles, ~5 nm in diameter, in contrast to other processes or synthesis routes where precipitation or reduction can produce colloidal (≤30 nm diameter) or larger nanoparticles (~50–100 nm).

Vapor, Spray, and Plasma Processes

Over the past two decades, various process technologies have been developed for producing a variety of nanomaterials or nanostructure forms: particles, films, or bulk-related (3D) nanomaterials by vapor condensation and especially rapid solidification, from atomized (vapor), aerosol, or liquid jets and sprays (Tjong and Chen 2004). Figures 1–7 illustrate a variety of these techniques, several of which are conceptually similar.

Figure 1 illustrates a variety of plasma synthesis techniques for producing nanoparticles. Figure 1a shows a flame which may conceptually be considered a partial plasma. In combustion flames such as propane ($C_2\ H_8$) or natural gas (methane, CH_4), variations in oxygen or air flow can alter the production of carbon nanoparticles, including soots, fullerenes, or carbon nanotubes. These can be collected electrostatically. Figure 2 illustrates multiwall carbon nanotube–multiwall (or multiconcentric) fullerene nanoparticle aggregates collected above a propane and methane (natural gas) flame, respectively, and observed in the TEM. Precursor materials, mixtures, or premixed gases can also be used to synthesize nanoparticles, for example, titania can be synthesized in a flame from premixed $Ti\ Cl_4 + CH_4 + O_2$. Using an electrostatic plate collection system not only collects $Ti\ O_2$ product, but the product is less aggregated and primary particles are smaller, nanoscale particles.

High-temperature plasma processes can use DC, AC, or RF power sources for nanopowder production using either liquid or solid powder precursors. Figure 1b shows two plasma synthesis arrangements. In the axial feed precursor on the left, a sheath gas draws the precursor into the system. Feeding the precursor into the plasma (flame) as illustrated in the right schematic in Fig. 1b is used for powder precursors. For RF plasmas, the nozzle design becomes crucial. Rapid cooling of

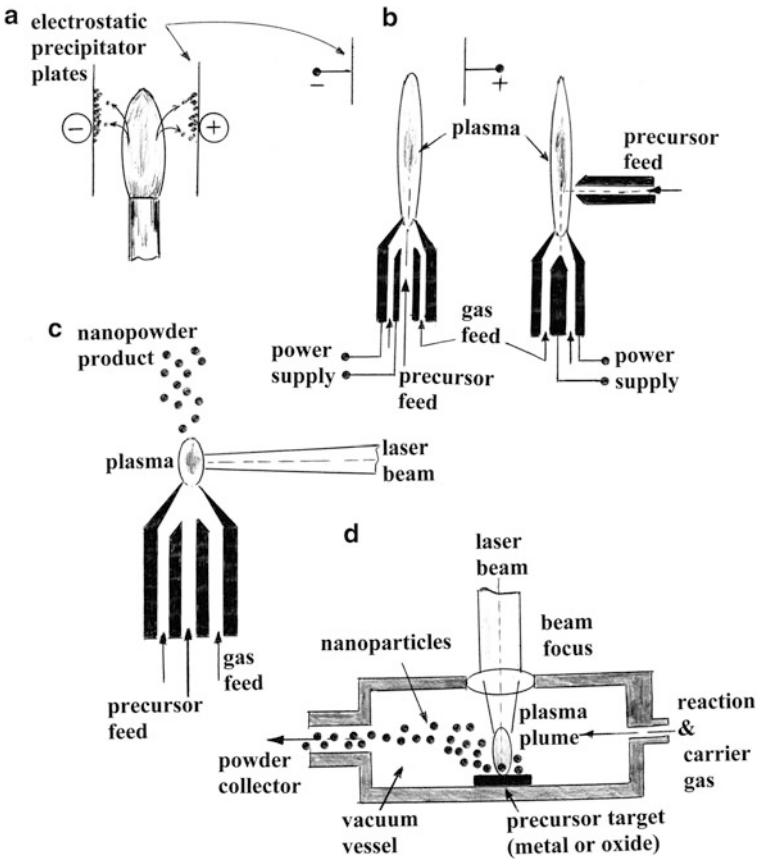


Fig. 1 Schematic views of plasma processes for nanoparticle synthesis. (**a**) Simple combustion flame with electrostatic collection plates. (**b**) Two basic and most plasma burners operating at atmospheric pressure using DC or AC power. Plasma burns between two coaxial electrodes, which are cooled by gas flow as well. (**c**) Laser-driven plasma system. (**d**) Laser ablation process. Reaction product is carried to cold surface or electrostatic collectors

the gas following reaction in the plasma is also important in controlling the nanoparticle size and size distribution as well as the degree of aggregation. In this respect, it is possible to project the plasma-generated nanoparticle stream onto a rotating, liquid nitrogen drum instead of collecting the product stream on the electrostatic precipitator plates.

In Fig. 1c the plasma is created by a focused laser beam. Reactant concentration is normally high since the focal point is just above the orifice, and this situation often creates agglomerated nanoparticles. Figure 1d utilizes laser ablation where the beam is focused onto a metal, alloy, or metal oxide target which is vaporized instantaneously, preserving the stoichiometry of the target. The supersonic, ionized vapor jet or plume of evaporated material in Fig. 1d can reach temperatures

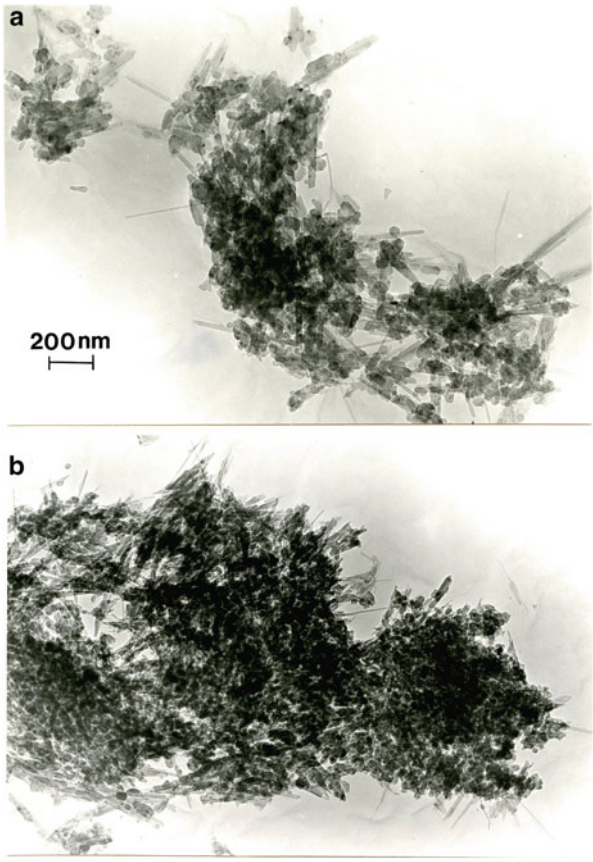


Fig. 2 Flame-produced multiwall carbon nanotube–fullerene aggregates observed in the TEM. (**a**) Propane flame. (**b**) Methane (natural gas) flame. Natural gas is 96 % methane

exceeding 3,800 K. Particle growth in the cooling, supersaturated plume is limited by system gas pressure. Increasing the supersaturation increases the number of nuclei and correspondingly decreases nanoparticle size in the powder product which can be collected electrostatically or on a cooled, rotating drum.

Cooled, rotating drum collectors originated in the 1960s when amorphous or glassy metal (or alloy) products were formed by supercooling on a liquid nitrogen cooled, rotating drum as illustrated schematically in Fig. 3a. In this process, liquid metals or alloys are projected onto the cooled drum, often referred to as a casting wheel. This process produced thin, even controlled thicknesses of amorphous ribbon which was rapidly stripped from the drum, often 3–5 cm in width and $<$100 μm thick, corresponding to cooling rates ranging from $\sim 10^6$ C° to 10^7 C°/s. These amorphous strips could be flash annealed to produce a brittle, nanocrystalline product which could be easily crushed to the nanograin size (Lu 1996). This process (Fig. 3a) was modified slightly as shown schematically in Fig. 3b to produce an aerosol precursor by mixing Ar gas with the jetting liquid. This produced a nanoparticle product which was scrapped from the cold drum.

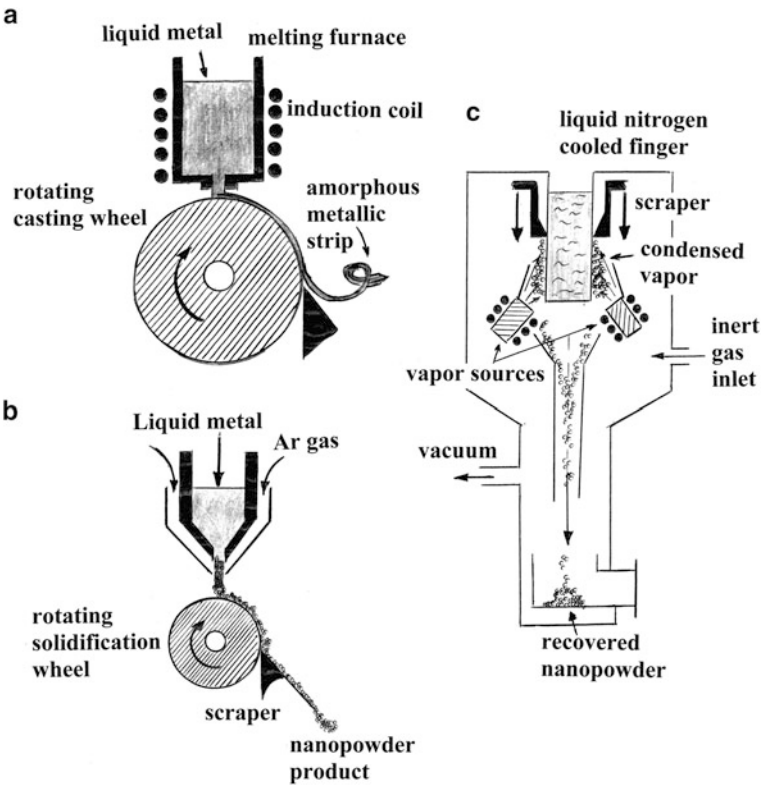


Fig. 3 Supercooling/rapid solidification processes. (**a**) Supercooled rotating casting drum process for producing amorphous metal or alloy ribbon. (**b**) Supercooled rotating drum process for producing nanoparticle product. (**c**) Evaporation–condensation process for nanoparticle (nanocrystalline) product synthesis. Evaporation sources are shown at S. Thermal evaporation is the most common source process

Figure 3c illustrates a later modification where vaporized metals or alloys were collected onto an extended cold finger or liquid nitrogen cooled cylinder fitted with a scraper assembly to systematically scrape the condensed, nanometal, or alloy product into a collection funnel, all in vacuum or an inert gas. This collection system in some cases also employed an assembly for consolidation of the powder using a piston and anvil assembly (Tjong and Chen 2004). Consolidated powder forms can then be placed in high-pressure piston–anvil assemblies to create nearly full density precursor billets for bulk nanomaterial fabrication.

Figure 4 compares common forms of physical vapor deposition (PVD) or chemical vapor deposition (CVD). As implicit in the designations, PVD is a physical (evaporation) process, while CVD is a chemical (reaction) process. As shown in Fig. 4a, a high-voltage or high-electric-field condition can create a plasma in an inert gas environment which strips electrons from the gas atoms and creates positive ions with an energy related to the electric field intensity. These ions interact

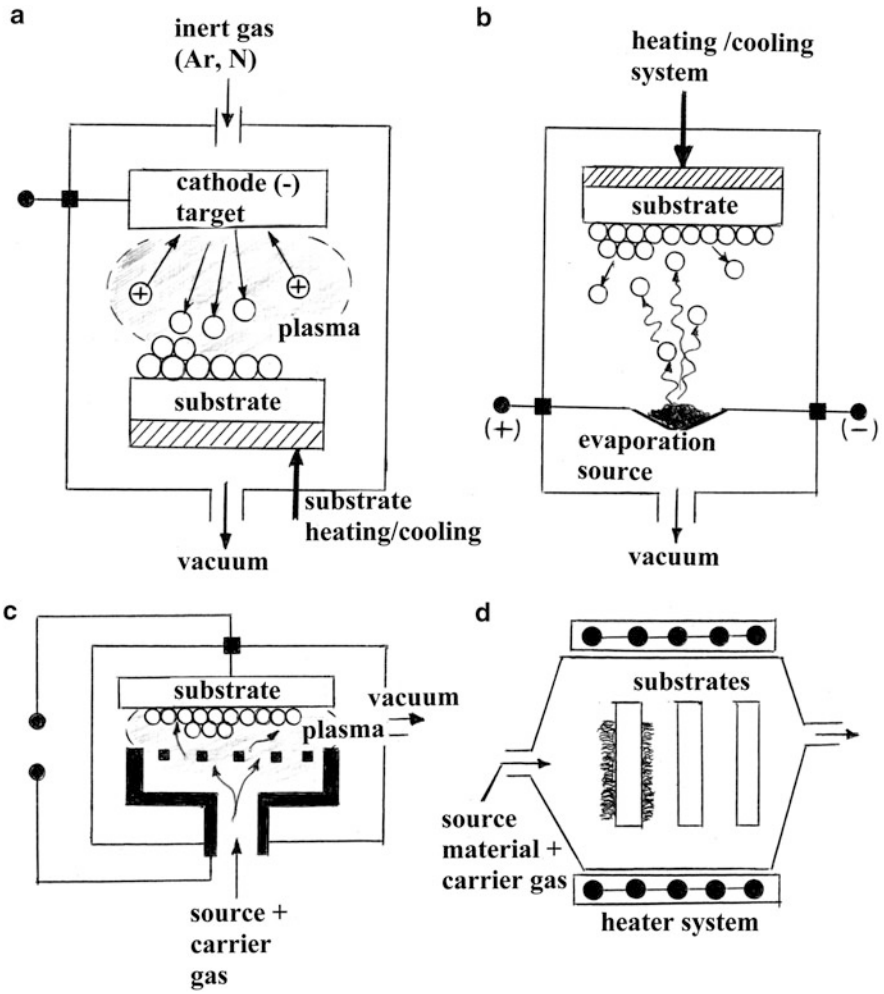


Fig. 4 Comparison of common physical vapor deposition (*PVD*) and chemical vapor deposition (*CVD*) process schematics. (**a**) Ion sputtering onto heated or cooled substrate. (**b**) Thermal evaporation and condensation of vapor on heated or cooled substrate. (**c**) Plasma CVD process where reaction product is deposited onto a substrate. (**d**) Thermal CVD process having multiple substrates

with (or impact) the cathode target where the target atoms are sputtered from the surface and directed toward the substrate where they condense, forming discontinuous islands or continuous thin films, depending upon the sputter energy, sputter time, and substrate temperature. Lower substrate temperatures favor nanostructures, while higher temperatures favor larger, continuous grain structures or even single-crystal films which can be nanometers in thickness. The structure of

the substrate (crystalline or noncrystalline) can also influence the sputter-deposited product.

Another more common form of PVD involves metal or alloy evaporation or even multiple source evaporation to create some required stoichiometry on the substrate surface. In this process, shown in the simple schematic of Fig. 4b, a material is evaporated from a source configuration using a number of forms of evaporative heating, the simplest of which is resistive heating in a "boat." This process is accomplished in a vacuum, often a very high (10^{-5} Torr) or ultrahigh ($>10^{-9}$ Torr) vacuum. The evaporating atoms are deposited on a substrate, either crystalline or noncrystalline, which may be heated or cooled. These parameters, along with the rate of evaporation, can be used to control the structure and continuity of the deposit. Deposits can become single-crystal films in contrast to nearly amorphous or nanograin films. The deposition process itself can involve a number of phenomena: adsorption, desorption, deposition, diffusion, nucleation, and growth. These will be treated briefly in Fig. 7b following.

In CVD processes, precursor components or some materials and a carrier gas, which can include water vapor, are mixed or injected into a plasma plume and reacted to form a prescribed nanomaterial which deposits on a substrate. Alternatively, the reaction proceeds in a thermal environment where the reaction product is deposited on a substrate or multiple substrates. These two processes are illustrated in Fig. 4c, d, respectively. It is apparent that the plasma CVD process in Fig. 7c is conceptually very similar to the plasma deposition processes of Fig. 1b. CVD processes (Fig. 4c, d) allow both thin films and nanoparticulates (nanopowders) to be fabricated, often in multiple substrate arrays as shown schematically in Fig. 4d. In most conventional CVD processing thin films form on the substrate, but process parameter adjustments can allow nanoparticle deposits instead of films. This form of deposition is also called *chemical vapor synthesis* (CVS). Precursor materials can be metalorganics (where the process is referred to as metalorganic CVD or MOCVD), carbonyls, hydrides, chlorides, or other volatile compounds in gaseous-, liquid-, or solid-state forms. Reactions, as illustrated, can be laser-assisted, microwave, or RF plasma processes.

CVD processes using hydrocarbon and especially hydrocarbon gas precursors are used to produce carbon nanotubes and diamond films as well as graphene, usually in high-temperature reactors or the so-called tungsten hot-wire reactors illustrated schematically in Fig. 5. Figure 5a employs metal nanoparticle catalysts either on substrate arrays or in some nanoparticle spray form which nucleates multiwall carbon nanotubes (MWCNTs). Simple mechanisms for this process are illustrated schematically in Fig. 5b, while Fig. 6b shows metal-catalyst-formed MWCNTs in comparison with graphite (or carbon) *arc-evaporation-produced MWCNTs* in Fig. 6a. In Fig. 5c, methane and hydrogen are injected into the system where the H_2 striking the hot W wire breaks up into atomic hydrogen, H (or 2H), which decomposes the CH_4 on the substrate surface forming either sp^3 carbon (diamond) or sp^2 graphite (graphene) hybridization, respectively, and hydrogen gas, depending on the CVD parameters. Graphene is usually deposited onto Cu, Ni, or Si

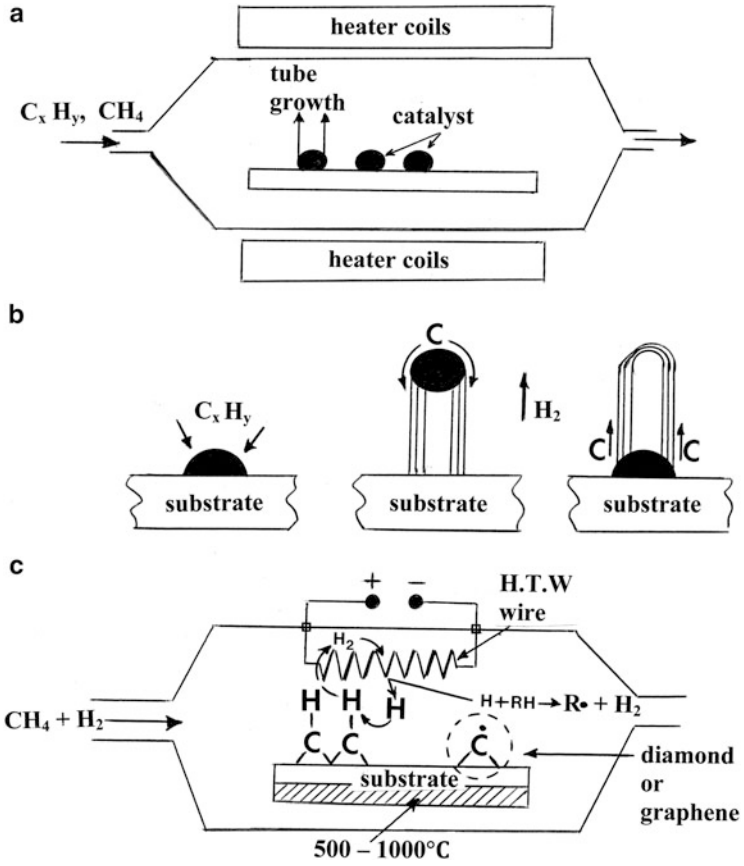


Fig. 5 CVD metal-catalyst-assisted growth of carbon filaments and multiwall carbon nanotubes (*MWCNTs*) from hydrocarbon precursors, C_x H_y ((**a**) and (**b**)). (**b**) Shows tip growth and base growth mechanisms for MWCNTs on metal catalyst nanoparticles. Adapted from Kumar and Ando (2010). (**c**) Growth of diamond films or graphene on heated substrate from methane precursor using a tungsten hot-wire process

substrates where it can be washed away and transferred to another desired substance surface.

In some cases, it is important to be able to fabricate multilayer nanocomposites, especially semiconductor heterojunction or related structures by the systematic deposition of nanolayers forming single-crystal films. This process is usually described in some form of *molecular beam epitaxy* (*MBE*), atomic layer epitaxy (ALE), or atomic layer deposition (ALD) as illustrated schematically in Fig. 7a. In this ultrahigh vacuum process, separate evaporation or effusion cells form a gaseous (vapor) element which is slowly deposited onto a single-crystal substrate. Often, co-deposition or multiple-element deposition of different elements occurs where these elements in predetermined stoichiometries react and grow epitaxially.

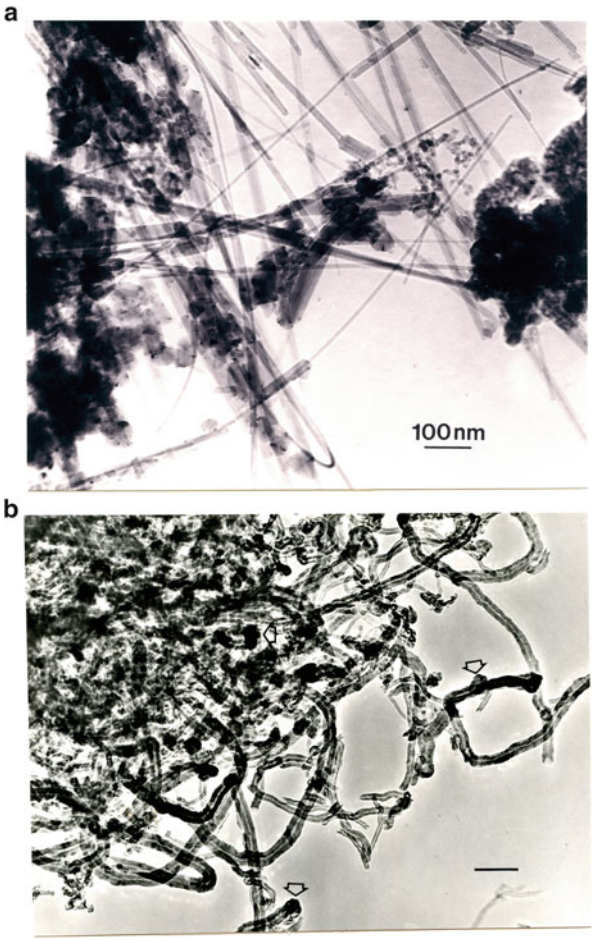


Fig. 6 Comparison of multiwall carbon nanotube production by carbon arc evaporation (**a**) and CVD catalysis (**b**). Arrows show metal (Ni) nanoparticles at tube ends. (**a**) and (**b**) are TEM images

For example, Al, Ga, and As can form predetermined layers of Al Ga As or Al As on Ga As and alternating in specifically defined nanothickness layers. Epitaxy occurs when the growing layers or films assume the crystal structure of the substrate. If the growth layer and substrate do not match in terms of unit cell dimensions, a mismatch occurs which produces interfacial strain. Such strained-layer device structures or strained-layer superstructures can have unique microelectronic and optoelectronic properties which are described in Part XV. Such structures are often referred to as quantum well or strained multiple-quantum well structures. Figure 8 illustrates a high-resolution TEM image of alternating Ga As and Al As layers. These structures are also referred to as atomic layer structures or superlattice structures. In some cases, deposited layers perfectly match the substrate crystal structure even though their lattice parameters do not match. As noted, this causes a strained lattice, and the process is called *pseudomorphic growth*. At some thickness, these layers may revert to the unit cell

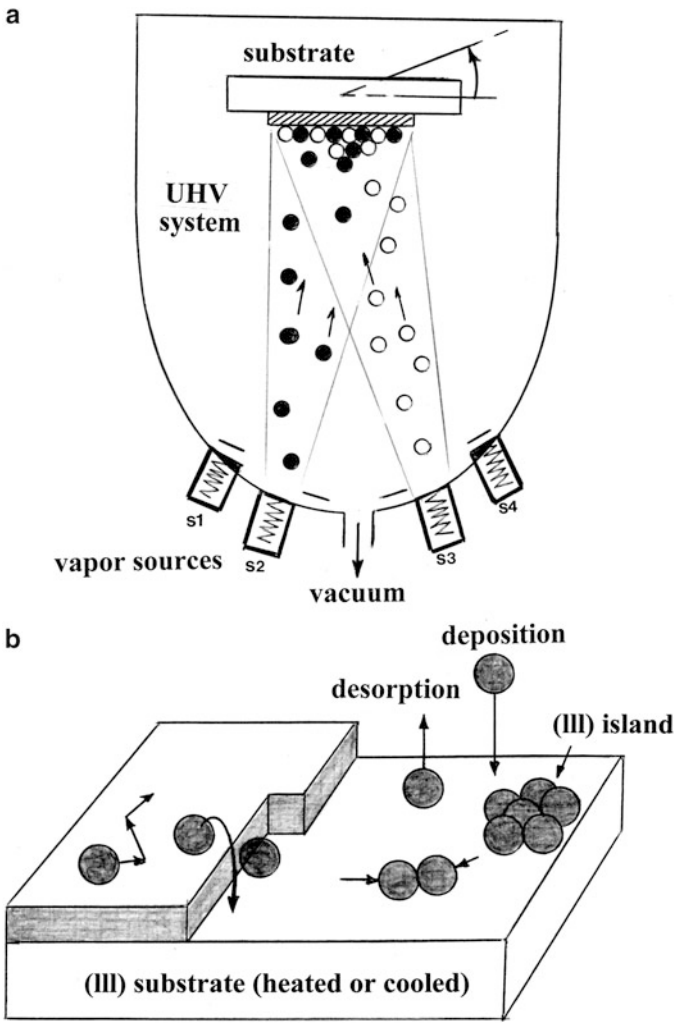


Fig. 7 MBE process schematic (**a**). Effusion (vapor) sources are noted 1, 2, 3, 4. Each source has a shutter arrangement. The substrate can be tilted or translated as necessary. It can also be heated or cooled. (**b**) Shows a cartoon depicting the overall deposition process on a crystalline substrate. (111) layer growing on (111) substrate. Diffusion of atoms on the surface and across and along growing edges is shown

dimension for the structure but still maintaining the orientation of the substrate in *epitaxial growth*. This transition is accommodated by dislocations forming at the interface.

The MBE process shown schematically in Fig. 7a is employed in electronic component/device fabrication where selective masking is used to selectively create device structures. These features will be described in more detail in

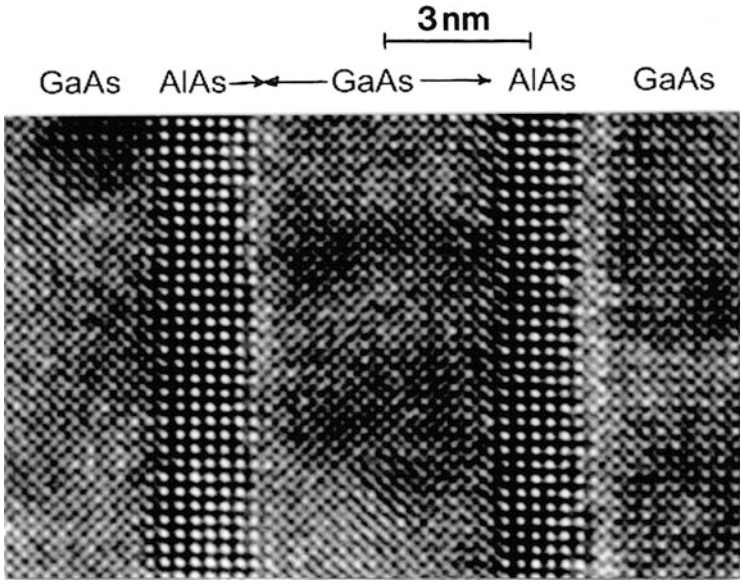


Fig. 8 High-resolution TEM image of alternating Ga As and Al As MBE-deposited layers in cross section (Courtesy of Simone Montanari, University of Liverpool)

Part XV. Masking can also be used to create periodic arrays of quantum dots and similar nanoparticle arrays on a substrate.

Figure 7b illustrates various phenomena involved in the deposition of atom species on a substrate and the growth of islands or thin films. In pseudomorphic or epitaxial growth, the island or film structure will be characteristic of the substrate crystal surface structure. Deposited atoms can, under some circumstance, desorb or leave the surface. For energetic (ballistic) atoms striking the surface, they may simply rebound. In some processes, this is characterized by a sticking coefficient or probability which is related to the fraction of atoms sticking or depositing. A probability of one implies all atoms will stick. Atoms on the surface can diffuse as indicated in Fig. 8. Different surface plane configurations such as edges or ledges are more conducive to atoms being accommodated in comparison to atoms which simply reside alone on the surface. This has to do with the energetics associated with unsatisfied bonds: ideally (1) for a single atom on the surface, (2) for an atom at a layer edge, and (3) for an atom at a ledge (or kink) in the layer. Nucleating atoms can cluster together and take advantage of this bonding energetics. In contrast to PVD processes or CVD processes, MBE reaction occurs at the substrate or in the vicinity of the substrate and not in the vapor stream. In effect, the substrate catalyzes the reaction, and correspondingly the surface orientation influences the energetics associated with such catalysis in the same way it does where catalysts are applied, especially nanoparticle catalysts exposing different surface facets or crystal orientations. In the case of catalysis, the rate of

reaction is significantly influenced. Or the propensity of adsorption of a species is influenced.

Considering the crude model of Cao (2004), the surface energy associated with a crystal orientation will be given generally by

$$\gamma = N_b \in_b \rho_a / 2 \tag{1}$$

where N_b is the number of broken bonds, $\in_b$ is the bond strength, and ρ_a is the surface atomic density or number of surface atoms per unit area on a newly created surface. For an fcc metal, for example, each surface atom on {100} facets would have four broken chemical bonds and Eq. 1 would become

$$\gamma(100) = 4 \in_b / a^2 \tag{2}$$

where a is the lattice parameter for the fcc unit cell. Similarly, {110} and {111} facets would have five and three broken bonds, respectively:

$$\gamma(110) = 5\varepsilon_b/\sqrt{2}a^2 \cong 3.54\varepsilon_b/a^2, \tag{3}$$

$$\gamma(111) = 2\sqrt{3}\varepsilon_b/a^2 \cong 3.46\varepsilon_b/a^2, \tag{4}$$

consequently γ (110) > γ (100). Nanocrystals with high surface energy have an open surface structure and a high density of low-coordinated step and kink atoms (Fig. 7b). Such features usually lead to exceptional catalytic properties. However, fundamental thermodynamics (the second law) forms lower-energy surfaces (or facets composing nanocrystals) (Zhou et al. 2011).

Catalysts reduce the activation energy barrier for reactions, and surface structure and energy make a major contribution to lowering this energy. However, the mechanisms involved in such reactions and product deposition on crystal surfaces is complex since deposition on a surface alters the structure and corresponding surface energy. In addition, islands formed create edges, ledges, and corner structures (Fig. 7b) all with increasing energy associated with them. Figure 9 illustrates crystal structure facets typical of nanocrystalline particles and catalyst particles and the fundamental deposition cluster appearance on (100) and (111) facet orientations. Figure 10 shows a ball model view of what might be considered an epitaxial (100) quantum dot or related single-crystal nanoparticle nucleating in the (100) orientation, characteristic of the 4-atom nucleus on (100) in Fig. 9 (see Scott et al. 2003).

Electrodeposition

Nanocrystalline deposits can be formed on a cathodic substrate in an ionic (DC) transport solution through a simple reaction of the form

$$M^{2+} + 2\overline{e} \rightarrow M^{\circ},$$

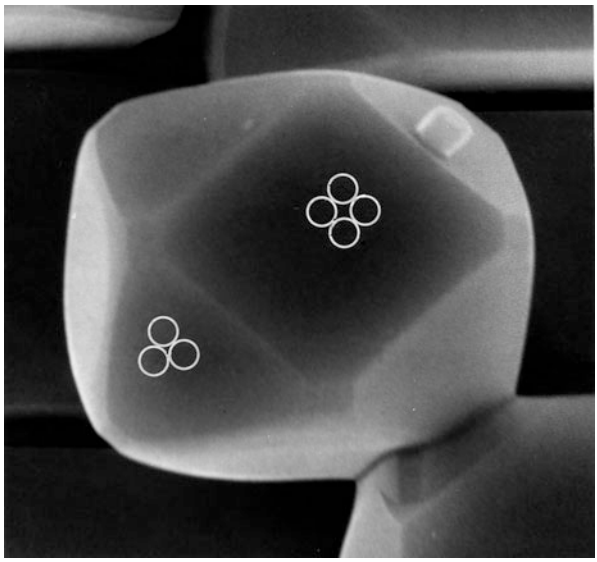


Fig. 9 Faceted MgO (NaCl crystal structure) nanocrystal observed in the SEM. Primary cuboctahedron with small (110) facets at the (100) plane corners. Corresponding unit clusters on (111) and (100) plane facets are shown

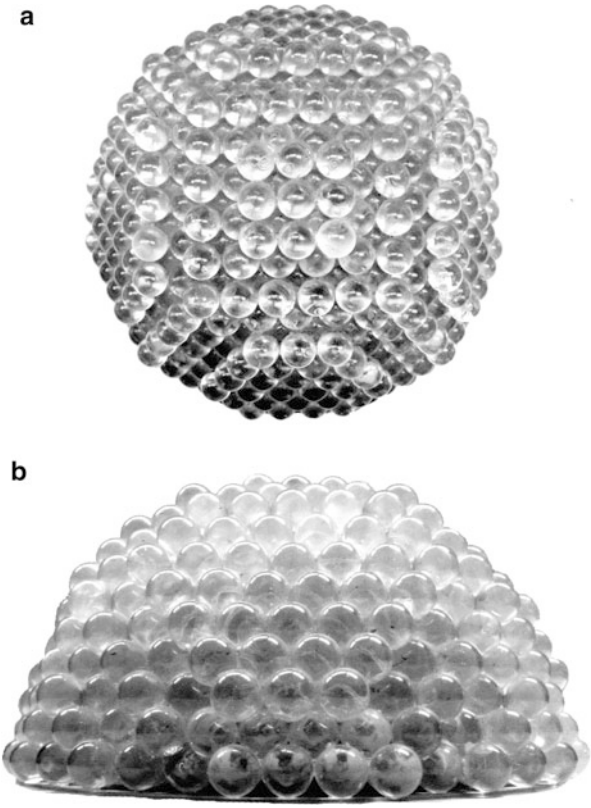


Fig. 10 [100] oriented ball model fcc crystal facet nanoparticle. (**a**) View down [100] axis. (**b**) Side view of nanoparticle model. Particle diameter is ~13 atom diameters or ~25 nm. True quantum dots would be roughly one fifth of this size or smaller. Note correspondence of (100) and {111} planes in (**a**) with Fig. 9

where M is a metal. The nanocrystal size and often nanostructure can be controlled by adjustment of bath composition, temperature, PH, current/voltage conditions, etc. Very fine grain or nanograin structure of electrodeposited coatings is often assured by using electrical pulses (*pulse plating*) where the pulse shape (current/voltage and time) can contribute to controlling the deposition parameters.

Notable examples of pulse plating have involved Ni (Robertson et al. 1999) and Cu (Lu et al. 2004). In the case of Ni, an equiaxed grain size of ~10 nm could be achieved where the ultimate tensile strength (UTS) was measured to be >2 GPa in contrast to 10 μm diameter grain size Ni electrodeposits where the UTS was 0.4 GPa. The corresponding Vickers microindentation hardness was 6.5 and 1.4 GPa, respectively (Robertson et al. 1999). Similarly, for electrodeposited nanocopper with a grain size of ~30 nm, the UTS was measured to be ~1 GPa in contrast to 100 μm grain size copper where the UTS was ~0.2 GPa (Lu et al. 2004). In either case, the difference in UTS or microindentation hardness was roughly a factor of 5. In the case of Cu, nanotwins (growth twins) also contributed to the microstructure development since the twin (or twin boundary) interfacial free energy is nearly half for Cu in contrast to Ni (see Part V, chapter "► Planar Defects: Crystal Interfaces").

Fabrication of Bulk Nanocrystalline Materials by Severe Plastic Deformation Processes

Nanoparticulate materials, even amorphous particulates fabricated as illustrated in Fig. 3, can be consolidated using a high-pressure (static) punch and anvil (or die) arrangement or by explosive (shock) consolidation which can apply very large dynamic pressure at relatively low temperatures ($\leq 100\ ^\circ$ C) (Murr 1988). Consolidation of cylindrical forms which can be used as billets as precursors for extrusions or other forming processes can produce nearly fully dense material.

In related, severe (plastic) deformation processes, high strain and strain rate (including impact or ballistic conditions) can create grain refinement by *dynamic recrystallization* driven by localized, adiabatic strain heating generally expressed by

$$\Delta T \sim \in \times \dot{\in} \tag{5}$$

where ΔT is the temperature gradient, $\in$ is the strain, and $\dot{\in}$ is the associated deformation strain rate. Such localized temperature gradients can drive chemical reactions in mixtures of oxides, chlorides, etc. These processes are often referred to as *mechanochemical synthesis* which can be characterized by the following displacement reaction:

$$M_xC + yB \rightarrow xM + B_yC, \tag{6}$$

where M_xC and B are precursor materials (M designating a metal). These materials or materials mixtures can be agglomerates which are placed in a ball mill or attriter

where repeated impacting can drive both a reaction as shown in Eq. 6 and grain or crystal refinement by dynamic recrystallization implicit in Eq. 5. This process can also be conducted where metal or alloy aggregates are not chemically reacted but are plastically deformed to create a dynamically recrystallized regime involving nanograins which can flow into one another and mix in the solid state. This process creates a *mechanical alloy* by repeated deformation in a ball mill or attrition, creating alloyed or immiscible mixtures and nanocrystalline aggregates (see Zehetbauer and Zhu 2009). The aggregates can form larger clumps which can be statically or dynamically consolidated as described above to create precursor billets for conventional processing: extrusion, forging, rolling, etc. Some of these features will be discussed in more detail in Part XIII which follows.

There are two other novel, extreme plastic deformation or *severe plastic deformation* (*SPD*) processes which can create nanocrystal or nanograin structures by dynamic recrystallization. These include friction-stir processing (FSP), a modification of friction-stir welding to be described in detail in Part XIII, and equal-channel angular processing (or pressing) (ECAP). In the former FSP concept, a rotating tool is translated along the surface, or within some depth of the surface, of a usually flat plate metal or alloy having a more conventional grain structure ($\geq$100 μm). This rotating tool causes the material to dynamically recrystallize and flow in the solid state. That is, the dynamically recrystallized grains (usually in the higher nano-regime: $>$ 500 nm) will glide one over the other (grain boundary sliding). As illustrated in Fig. 11, this process can create a nanocrystalline layer on the surface of a conventionally processed plate, or a thin layer of some new material (metal or alloy) can be deposited on the surface of the plate, and then stir processed to create a nanocrystalline regime. This stir processing can be controlled by the length of the stirring tool as well as the rotation and translation speeds, which can alter the strain and strain rate as implicit in Eq. 5. The stir processing can raster across a selected surface area or process the entire surface, creating an engineered surface layer.

In the ECAP process (also referred to as *equal-channel angular extrusion*) illustrated schematically in Fig. 12, precursor powders or precursor consolidated billets can be further refined to produce very small nanocrystalline grain structures, or more conventionally processed metal or alloy billets having conventional grain sizes ~100 μm can be strained to create dynamic recrystallization producing nanocrystal grain structures.

Precursor billets are pressed through the die arrangement shown in Fig. 12 using a plunger. The die geometry, or angular relationships, governs the effective strain (ε) according to the relationship

$$\varepsilon = [2\cot(\phi/2 + \varphi/2) + \varphi\,\mathrm{cosec}\,(\phi/2 + \varphi/2)]/\sqrt{3}. \tag{7}$$

Figure 13 illustrates effective stress contours for a finite element analysis in the recent work of Nagasekhar et al. (2007) as the inside die angle, ϕ, was changed. Corresponding to larger effective stresses or strains in response to ECAP die angles implicit in Fig. 13 or Eq. 7, Figure 14 illustrates a few examples of the nanograin sizes achieved for different die angles (creating the effective strains noted as

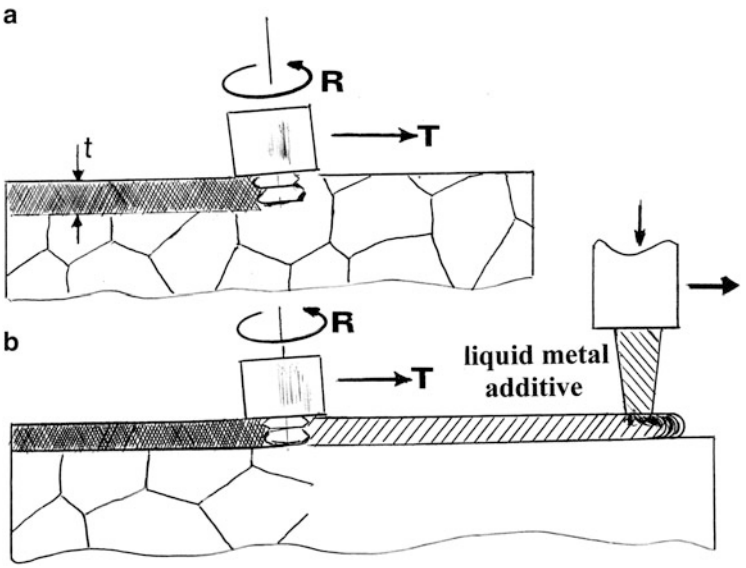


Fig. 11 Friction-stir processing of bulk plate surface (**a**) and additive surface layer (**b**). The additive layer can be laser-assisted direct metal deposition, laser-engineered net shape deposition, etc. See Part XI. R and T in (**a**) and (**b**) refer to the pin tool rotation (R) and traverse (T) speed. The tool is slightly off the normal axis as shown for efficient stir processing

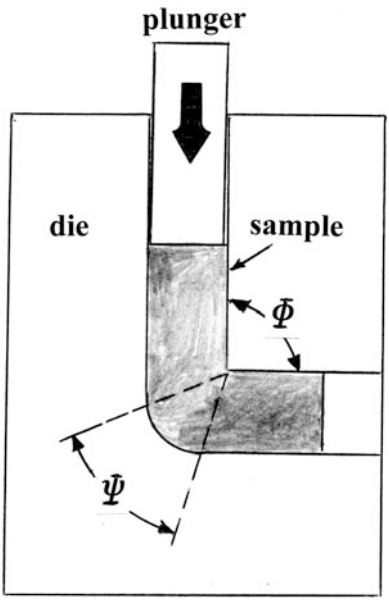


Fig. 12 Equal-channel angular processing (or pressing) (*ECAP*) for grain refinement

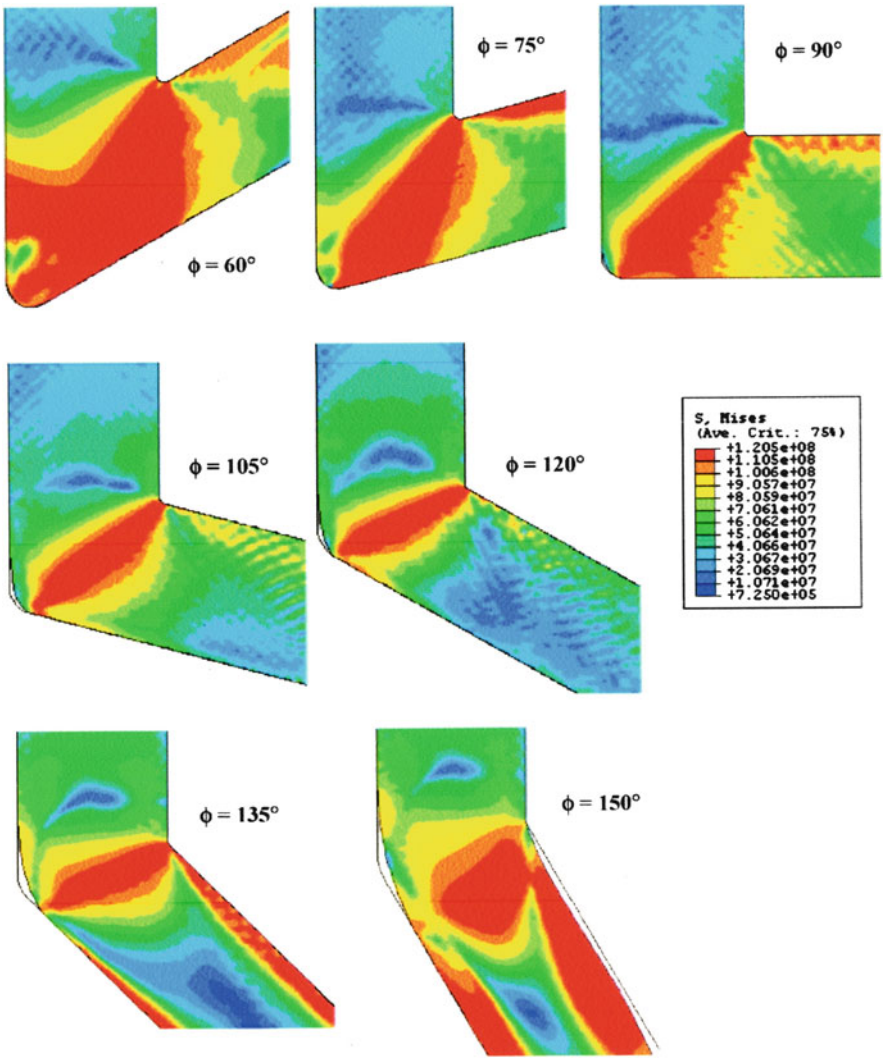


Fig. 13 Finite element simulations for effective stress contours related to different ECAP channel angles (Adapted from Nagasekhar et al. (2007))

expressed in Eq. 7), with a single pass of the precursor billet, as well as multiple passes at a constant 90° die angle. Strain is partly cumulative, but in addition, grains can be restructured by repeated dynamic recrystallization which can result in grain texture as well as increasing grain boundary misorientation (higher energy grain boundaries). Variations of the so-called process routes in ECAP can produce elongated grains (texture) as well as equiaxed grains implicit in Fig. 14c (Valiev and Longdon 2006).

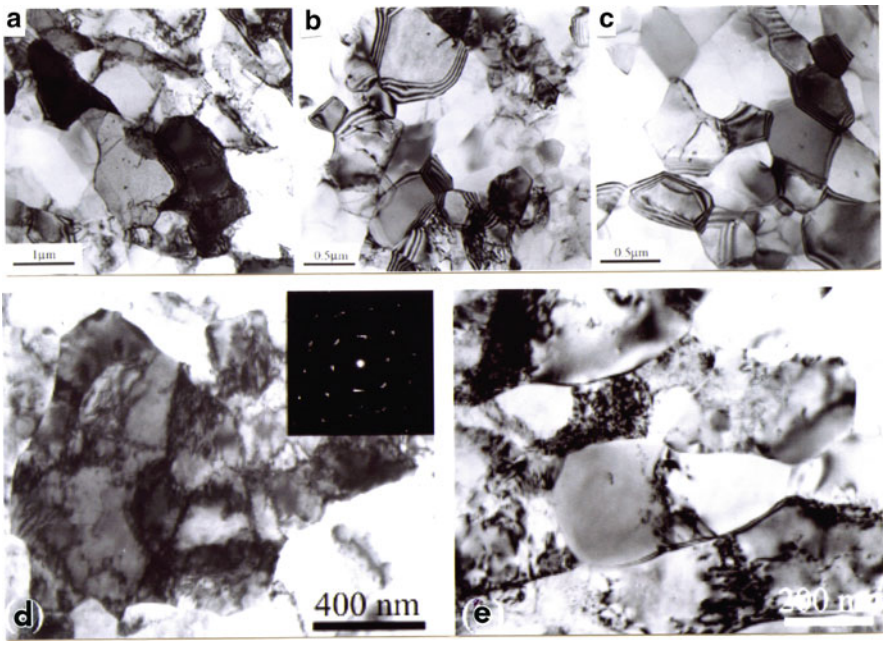


Fig. 14 TEM image examples of nanograin structures related to ECAP processing. (**a**) to (**c**) are Al processed at equivalent strains of 2, 4, and 8, respectively. (**d**, **e**) show Cu processed at an angle $\phi = 90°$; 4 and 24 passes, respectively. SAED pattern insert in (**d**) shows some arcing of spots due to small grain size. (**a**–**c**) (Courtesy of A.A. Kibar, Middle East Technical University; (**d**, **e**) are courtesy of C.X. Huang, Shenyang National Laboratory for Materials Science, Shenyang, China)

Bulk nanometals such as those processed by ECAP can exhibit superplasticity because the nanograins promote easy grain boundary sliding rather than more conventional, dislocation-driven deformation. These properties and performance issues will be discussed later.

References

Ahmad P, Mukherjee S, Senapati D, Mandal MI, Khan RK, Sastry M (2003) Extra cellular biosynthesis of silver nanoparticles using the fungus *Fusarium oxysporum*. Colloids Surf B Biointerfaces 28:313–318

Anastas PT, Warner JC (2000) Green chemistry: theory and practice. Oxford University Press, New York

Brinker CJ, Scherer GW (1990) Sol–gel science: the physics and chemistry of sol–gel processing. Academic, New York

Cao G (2004) Nanostructures and nanomaterials: synthesis properties and applications. Imperial college Press, London

Das SK, Marsili E (2010) A green chemical approach for the synthesis of gold nanoparticles: Characterization and mechanistic aspect. Rev Environ Sci Biotechnol. Published online 4 Feb 2010. doi:10.1007/S11157-010-9188-5

Kumar M, Ando Y (2010) Chemical vapor deposition of carbon nanotubes: a review on growth mechanisms and mass production. J Nanosci Nanotechnol 10:3739–3758
Lu K (1996) Nanocrystalline metals crystallized from amorphous solids: nanocrystallization, structure and properties. Mater Sci Eng R16:161–221
Lu L, Shen Y, Chen X, Qian L, Lu K (2004) Ultrahigh strength and high electrical conductivity in copper. Science 304:422–426
Marchiol L (2012) Synthesis of metal nanoparticles in living plants. Ital J Agron 7(e37):274–282
Murr LE (ed) (1988) Shock waves for industrial applications. Noyes, Park Ridge
Nagasekhar AV, Tick-Hon Y, Seow HP (2007) Deformation behavior and strain homogeneity in equal channel angular extrusion/pressing. J Mater Process Technol 192–193:449–452
Robertson A, Erb U, Palumbo G (1999) Practical applications for electrodeposited nanocrystalline materials. Nanostruct Mater 12:1035–1040
Scott SL, Crudden CM, Jones CW (eds) (2003) Nanostructured catalysts, Nanostructure science and technology series. Springer, New York
Tjong SC, Chen H (2004) Nanocrystalline materials and coatings. Mater Sci Eng R 45:1–88
Valiev RZ, Longdon TG (2006) Principles of equal-channel angular pressing as a processing tool for grain refinement. Progress in Materials Science, vol 51. Issue 7, Elsevier, New York
Zehetbauer MJ, Zhu YT (eds) (2009) Bulk nanostructured materials. Wiley-VCH, Weinheim
Zhao P, Li N, Astroe D (2013) State of the art in gold nanoparticle synthesis. Coordin Chem Rev 257:638–665
Zhou Z-Y, Tian N, Li J-T, Broadwell I, Sun S-G (2011) Nanomaterials of high surface energy with exceptional properties in catalysis and energy storage. Chem Soc Rev 40:4167–4185

Mechanical and Other Properties of Nanocrystalline Materials

Contents

Abstract

Properties, especially mechanical properties of nanocrystalline metals and alloys are described along with other behavior related to nanocrystal structures which create quantum confinement at dimensions <6 nm. The transition from the larger length scales in the characterization of engineering polycrystalline materials through the nanoregime to the amorphous state is discussed in relation to polycrystal grain size D in relation to the grain boundary or grain boundary phase width, Δt: $D >> \Delta t$ (coarse-grain, micron regime), $D > \Delta t$ (nanograin regime), and $D \sim \Delta t$ (amorphous phase regime). Over these length scales, a Hall–Petch relationship is shown to be generally applicable up to a plateau which converges to the amorphous state in common metals such as Cu (fcc) and Fe (bcc). Deformation mechanisms also change from slip (or dislocation generation, glide, and pileup) to grain boundary sliding, to mixtures of these. In nanocrystalline metals, stacking-fault and twin boundary free energies are suppressed and partial dislocation emission and associated twin formation can occur. Deformation in small samples can favor surface sources in contrast to grain boundary dislocation sources depending upon grain boundary structure and the number of grains in the specimen thickness. There is a critical specimen thickness, Δ-to-grain size, D, ratio for nanocrystalline metals similar to engineering test samples having coarse-grain sizes, where $\Delta/D > 8$.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_47

Introduction

Chapter "► Classifications and Structures of Nanomaterials" outlined some simple classifications relating to nanoparticulate materials in the context of free nanoparticles, nanoparticulates in a solid-state regime, especially a crystalline regime or a polycrystalline regime where the included nanoparticle length scales may be compared to the crystal or grain size length scales. The latter includes polycrystalline solids where the grain or phase size may span length scales from the macro (mm) through the micro (μm) and nano (nm) to the amorphous state as implicit in ► Fig. 21a–f of Chapter "Classifications and Structures of Nanomaterials." Also implicit on considering so-called free particle size is the associated length scale which ideally might be considered to range from the m-to-cm regime to the nm regime in the context of mechanical testing, even hardness testing as this in itself may apply to the length scale regimes: Rockwell (mm) indenters to nanoindenters (nm). All of these issues influence the nanoscale behavior or properties of materials, including mechanical, electrical, optical, etc., as well as the performance and applications of nanomaterials. The latter will also include the effects of nanomaterials, especially free nanoparticles or nanoparticle aggregates on public health, i.e., human toxicity, which will be addressed in Chapter "► Performance, Applications and Health Concerns of Nanomaterials" following.

In the case of nanoparticles or discrete nanostructures having periodic nano dimensions (such as the superlattice structure shown in ► Fig. 8 in chapter "Synthesis and Processing of Nanomaterials"), quantum effects or quantum structures in which quantum effects become dominant are those where the length scale must be smaller than the de Broglie wavelength:

$$\lambda = h/P, \tag{1}$$

where, for the electron, P (the momentum) = $hk/2\pi$ (where h is Planck's constant and k is the Boltzmann constant. Taking the kinetic energy to be equal to (3/2) kT = $P^2/2m_o$, where m_o is the electron rest mass, and assuming room temperature (T ~ 300 K) and substituting for h and k, the value of λ in Eq. 1 becomes ~6 nm. Consequently, quantum confinement occurs when the nanostructure size is below 6 nm. This includes semiconducting quantum dots or superlattice structures illustrated in ► Fig. 8 of chapter "Synthesis and Processing of Nanomaterials." Like many confining structures, quantum dots can resemble an artificial atom containing, or are susceptible to, definable excited state transitions as a consequence of the ability to manipulate their energy levels. MBE systems can fabricate patterned quantum dots on specific substrates, especially crystalline substrates or they can be self-assembled. Quantum dot devices can be fabricated by growing quantum dots on a semiconductor to create a p-i-n laser diode: red-emitting quantum dots are obtained using highly strained (pseudomorphic) InAlAs in AlGaAs barriers on a GaAs substrate. Tunable spectra are also possible, and InAs/GaAs quantum dot infrared photodetectors can be fabricated in semiconductor

layer structures as shown in ▶ Fig. 8 of chapter "Synthesis and Processing of Nanomaterials." Photonic crystals with tunable properties can also be fabricated as larger nanostructures implicit in ▶ Fig. 10 in Chapter "Synthesis and Processing of Nanomaterials." These features will be discussed in more detail in Part XV.

In bulk nanocrystalline materials, it is also of interest to examine the effects of confinement as this may influence such phenomena as dislocation emission from sources, especially grain boundary ledge sources, or even the source structure, as well as slip and interaction of dislocations in a nanoconfinement. Of course, as implicit in ▶ Fig. 21a–f in Chapter "Classifications and Structures of Nanomaterials," crystalline structure, especially nanostructure, will converge to the amorphous state, where dislocations do not exist in the sense of definition in the crystalline state (see ▶ Fig. 1 in Chapter "Line Defects: Dislocations in Crystalline Materials"). However, Li (1976) has defined the amorphous state as a dislocation lattice and slip in this dislocation lattice visualized as a dislocation in a dislocation lattice as illustrated in Fig. 1. Indeed, slip line features are observed in deformed metal or alloy glass materials similar to those observed in deformed polymers such as polystyrene. In addition, if polycrystalline metals or alloys are heavily deformed, it can be estimated that the structure would become essentially a dislocation lattice structure like ▶ Fig. 1b of chapter "Synthesis and Processing of Nanomaterials" when the dislocation density approaches $\sim 10^{14}/cm^2$. However, this dislocation density in real polycrystalline metals and alloys produces so much free energy volume increase that recrystallization or early stages of recrystallization occur just as strain softening in the stress–strain diagram. Consequently, the amorphous state is more easily achieved by going from the liquid phase to the solid phase by rapid cooling (or rapid solidification processing illustrated in ▶ Fig. 3a of chapter "Synthesis and Processing of Nanomaterials"). This is a metastable state which will also revert to a more stable crystalline (nanocrystalline) (or equilibrium) state at the so-called glass transition temperature, which for many glassy metals and alloys is ~450 °C.

While Li (1976) demonstrated that the shear modulus for glassy (amorphous) metals and alloys was ~0.6 that of crystalline metals and alloys, the strength and hardness of amorphous metals and alloys often exceeds that of crystalline metals and alloys by more than a factor of 5. Murr (1980) attempted to rationalize this by considering a form of the Hall–Petch equation:

$$\sigma = \sigma_O + K_1 D^{-1/2} + K_2 \Delta t^{-m}, \tag{2}$$

where K_1 and K_2 are presumably different constants which relate to the strength of the grain matrix or boundary regime, respectively, D is the grain size, and Δt is the grain boundary width, which as represented in ▶ Fig. 8f of chapter "Synthesis and Processing of Nanomaterials" for the amorphous state is equal to the grain size: $\Delta t = D$. The exponent m in Eq. 2 was assumed to vary between 0.25 and 1 but equal to 0.5 for $\Delta t = D$. Equation 2 is similar to one developed by Langford (1977) to express the yield strength of fine pearlite in the form

$$\sigma = \sigma_o + K_1 \lambda^{-1/2} + K_2 \lambda^{-1}, \tag{3}$$

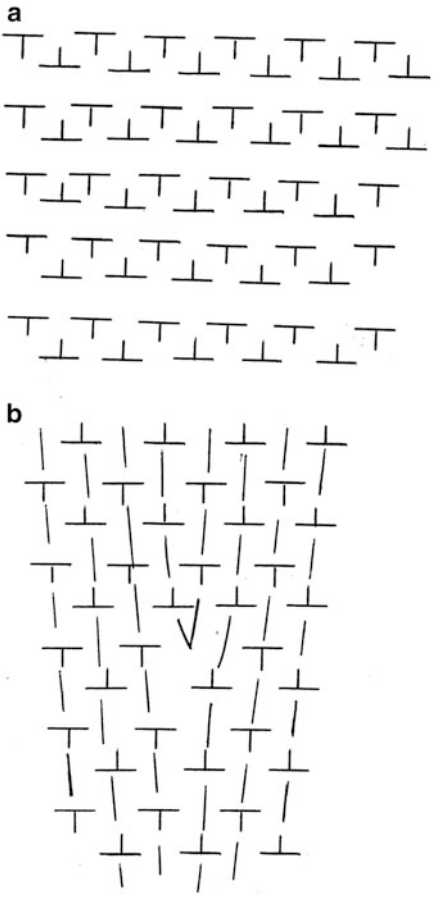


Fig. 1 Dislocation lattice model for an amorphous metal (**a**) and corresponding representation of edge dislocation in a dislocation lattice (**b**) (Adapted from Li (1976))

where λ is the interlamellar spacing. This derives from the concept of a deformation-gradient wavelength, λ, described earlier by Ashby (1970) in the general form

$$\sigma = \sigma_o + \sum K_s \lambda^{-m}, \tag{4}$$

where K_s is referred to as specific boundary strength constant.

By reconfiguring ▶ Fig. 20 of chapter "Classifications and Structures of Nanomaterials" and ▶ Fig. 8d and e of Chapter "Synthesis and Processing of Nanomaterials," a contrast is readily observed between more conventional engineering polycrystalline materials where the grain boundary width, Δt (estimated to vary from about 5 to 20 Å), is much less than the grain size, D: $\Delta t << D$ (and D ~ 100 μm), and the nanocrystalline regime, where $\Delta t \leq D$. This is illustrated in Fig. 2.

Thompson and Flanagan (1973) and Margolin and Stanescu (1975), among others, contended that the grain boundary regime (Δt) and the grain interior or

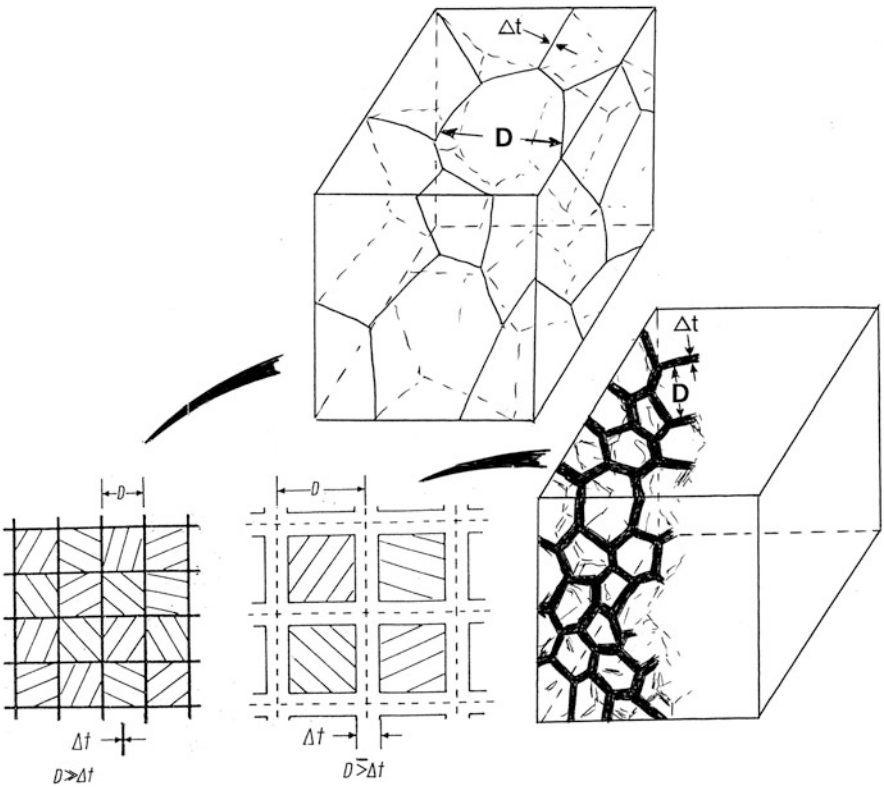


Fig. 2 Schematic comparison of conventional, polycrystalline engineering metal with a nanocrystalline metal. Grain size is represented by D and the grain boundary width or dimension is represented by Δt

matrix (D) have different flow stress. Margolin and Stanescu (1975) assumed that $\overline{\sigma}_{gb} > \overline{\sigma}_B$, where $\overline{\sigma}_{gb}$ is the average grain boundary flow stress and $\overline{\sigma}_B$ is the average grain or matrix flow stress. They based this assumption to some extent on their observation that slip band formation in a β-Ti alloy started at the grain boundaries and proceeded into the grain interior. In addition, they measured average volume fractions of grain boundary deformation and obtained average values for $\overline{\sigma}_{gb}$ and $\overline{\sigma}_B$ which showed that both increased with increasing strain, but at different rates. From these observations, they represented the total stress by

$$\sigma_T = \overline{\sigma}_{gb} V_{gb} + \overline{\sigma}_B V_B, \tag{5}$$

where V_{gb} and V_B are the grain boundary and grain interior volumes, respectively. Similarly, Murr and Wang (1982) illustrated experimentally that dislocation density in a region near the grain boundary in 304 stainless steel and nickel increased differently with strain and was greater initially at the grain boundaries

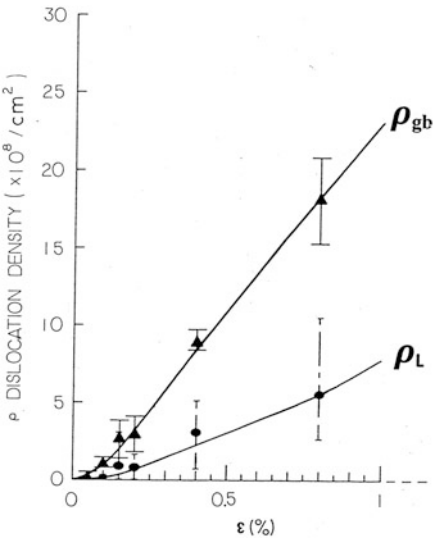


Fig. 3 Dislocation density measurements in deformed 304 stainless steel sheet strained in tension at low strain values, ε in %. ρ_{gb} is the average dislocation density in a region extending 1 μm on either side of grain boundaries observed in the TEM. The grain size (average) was 16 μm. ρ_L is the average dislocation density measured in the grain centers (Adapted from Murr and Wang (1982))

than in the grain interior. This is illustrated in data reproduced in Fig. 3 and in TEM observations shown in ▶ Figs. 18 of chapter "Line Defects: Dislocations in Crystalline Materials" and ▶ 34 of Chapter "Planar Defects: Crystal Interfaces" previously.

Figure 4 illustrates the implications implicit in Figs. 2 and 3 as these relate to nanocrystalline materials. Figure 4a shows a difference in the elastic and shear moduli (E_L, E_L', and E_{gb}; G_L, G_L', and G_{gb}) where different grain orientations may dictate different values for G_L and G_L' or E_L and E_L'. Dislocation structures in the grain boundary regime represent intrinsic grain boundary dislocations (⊥) and extrinsic grain boundary dislocations (⊥) which are ledges or ledge-related. Figure 4b represents the implications of the experimental data in Fig. 3 and similar data for nickel (Murr and Wang 1982).

Strength of Nanocrystalline Metals

It is of interest to examine the strength of nanocrystalline metals in the context of more contemporary engineering polycrystalline metals which are often described by a Hall–Petch equation of the form

$$\sigma_y = \sigma_o + K/\sqrt{D} \tag{6}$$

where σ_y, the yield stress, is also often related to the Vickers microindentation hardness (H_v) by

$$\sigma_y \cong H_v/3 \tag{7}$$

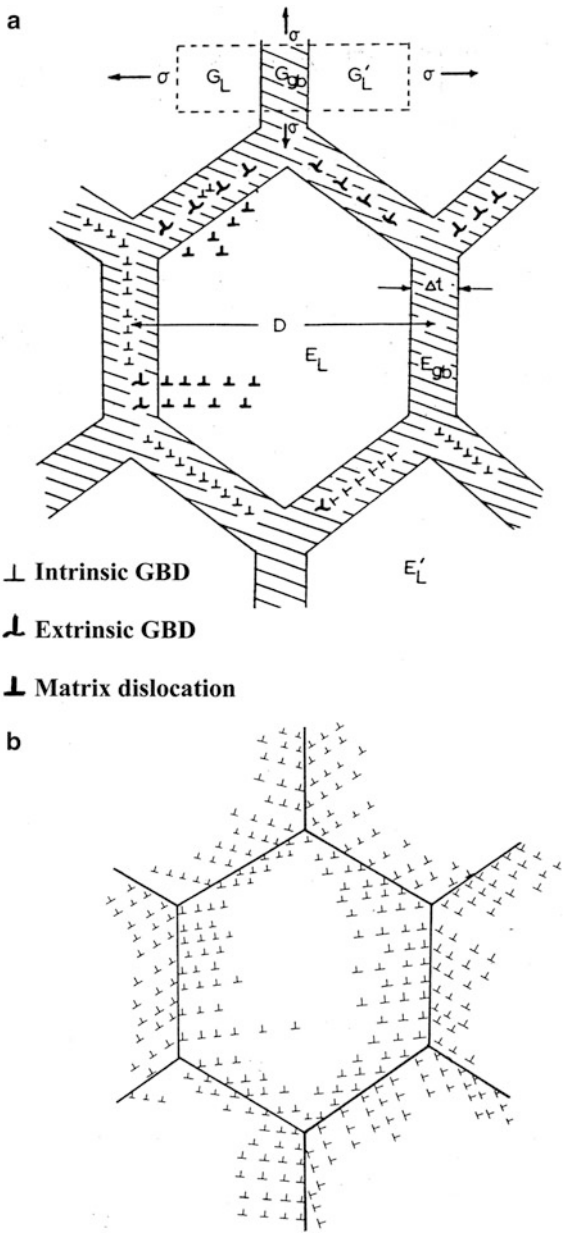


Fig. 4 Schematic illustration of grain structure (**a**) and dislocation generation and density near the grain boundary (**b**)

► Figure 6 in Part V, Chapter "Strengthening by Crystal Imperfections," provides a summary of metals and alloys which are characterized by Eq. 6, for grain sizes ranging from $>10^2$ to $\sim$10 μm. These more conventional (or commercial) engineering metals and alloys can range in bulk dimensions to tens of meters, and

the grain structures (or diameters) and their microstructures remain essentially constant, especially for applications where temperatures are near room temperature and loads (or stresses) do not exceed the yield stress. With the emergence of amorphous or glassy metals and alloys in the mid-to-later 1960s, their hardness and yield strength in comparison to polycrystalline structures, which ranged to values as high as 3 and ~1 GPa, respectively, were increased to 9 and ~3 GPa, respectively, with reference to Eq. 7 (Ashby and Greer 2006). However, the use of amorphous metals or metallic glasses as structural materials was limited as a consequence of their rapid solidification processing (▶ Fig. 3a of chapter "Synthesis and Processing of Nanomaterials") which produced only thin ribbon (~5 cm wide and 0.1 mm thick).

In the mid-1970s, Lashmore et al. (1977) prepared CVD α-Fe whiskers having grain sizes which were examined by TEM and FIM (▶ Fig. 10 of chapter "Classifications and Structures of Nanomaterials") to range in size from ~2 to 20 nm (Wilsdorf et al. 1978). Using Eq. 6, yield stresses were calculated to exceed 5 GPa, and tensile measurements on whiskers as large as 20 μm exhibited comparable values. However, the TEM and FIM observations illustrated that these whiskers consisted of a unique and complex nanodispersion of iron oxides, carbides, and atomic carbon bonded with the α-Fe nanograins to form a mixed covalent–metallic bonding which was responsible for the exceptional strength of the whiskers (Wilsdorf et al. 1978). Rack and Cohen (1976) also observed yield or flow stress values of ~1 GPa in steel wires cold drawn to true strains of up to 7, which resulted in grain and sub-grain diameters of ~300 nm. Gleiter (1981) utilized a rapid cooling or rapid solidification process as illustrated in ▶ Fig. 3c of chapter "Synthesis and Processing of Nanomaterials" to produce nanoparticulate metals and alloys which when recovered and consolidated into high-density solids exhibited elevations in hardness and yield strength according to Eqs. 6 and 7.

In spite of these early developments along with a wide range of studies over the past nearly half century, bulk nanocrystalline metals and alloys, as well as other nanocrystalline ceramics, etc., have not been produced in large dimensions, such as precursor billets and plates for rolling or extrusion, etc. This is due in part to the metastability of very small nanograin metals and alloys as recently discussed by Weertman (2012). Despite these limitations, an understanding of mechanisms underlying the deformation behavior of nanocrystalline metals and alloys has contributed to understanding extreme deformation processes such as machining, wire drawing, explosive welding, friction-stir welding and processing, ballistic impact, penetration and perforation, etc., to be described in more detail in Part XIII which follows.

To provide a conceptual context for considering nanocrystalline metals in comparison to more conventional, engineering polycrystalline metals, it is useful to look at the important issues regarding polycrystalline structure and how it contributes to strength and deformation behavior. Figure 5 shows a section representing Fig. 2. In this section grains are joined at triple junctions and grain boundaries make an angle θ with the reference surface or section surface. For grains creating a junction at A, B, and C in Fig. 5, where the orientation is the same, the

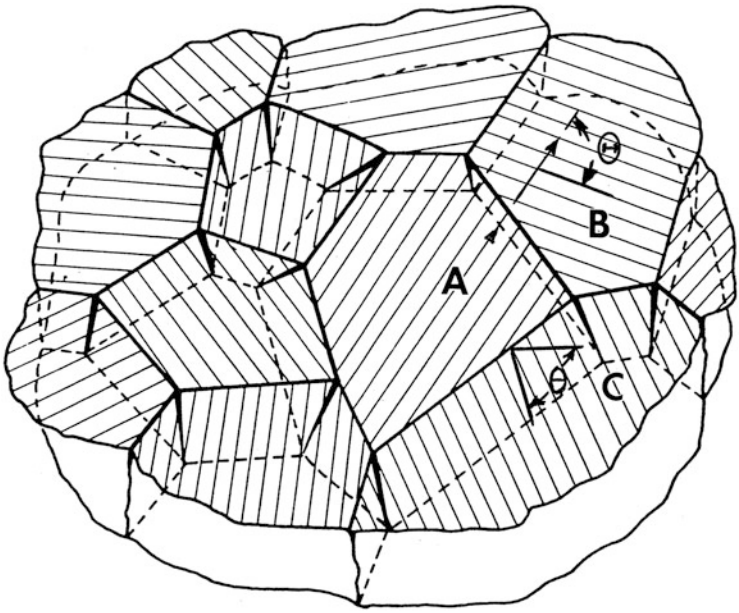


Fig. 5 Polycrystalline metal or alloy section in 3D solid. Grain boundary misorientation between grains A and B is denoted Θ. θ represents the inclination of the grain boundaries relative to the reference surface section

projection of a common crystallographic direction across the grain boundary represents the misorientation angle, Θ, which characterizes low-angle or low-energy boundaries relative to high-angle or high-energy boundaries as described in detail earlier in Part V, Chapter "► Planar Defects: Crystal Interfaces" (Fig. 17). In the ideal situation where $\theta = 90^\circ$ and grains A, B, and C in Fig. 5 have the same orientation ([hkl]), the misorientation, Θ can uniquely characterize the interfacial (or grain boundary) free energy which can be resolved along the interface as illustrated in a high-resolution, TEM lattice image for an aluminum triple junction shown in Fig. 6a. It can be observed in Fig. 6a that the misorientation across the grain boundary separating grains A and B is $\Theta_{AB} = 42^\circ$. This boundary is a symmetrical boundary since the misorientation angle, Θ_{AB}, is bisected by the boundary plane or interface plane (see ► Fig. 19b in Chapter "Planar Defects: Crystal Interfaces"). Correspondingly, the misorientation across AC and BC (which are asymmetric grain boundaries) is lower: $\Theta_{AC} = 25^\circ$; $\Theta_{BC} = 18^\circ$; and the associated grain boundary free energies resolved in the interface planes (or Gibbs dividing surfaces) as shown by the white arrows is less than that for the grain boundary separating grains A and B. It can be observed that the arrow resolved in AB pulls with more "force" or energy against the other two resolved arrows (across AC and BC). It can be noted that in these grains, the grain boundary width is only about two interplanar spacings wide. This is more readily observed on comparing grain boundary images shown in Fig. 6b and c. It should be noted in Fig. 6 that the

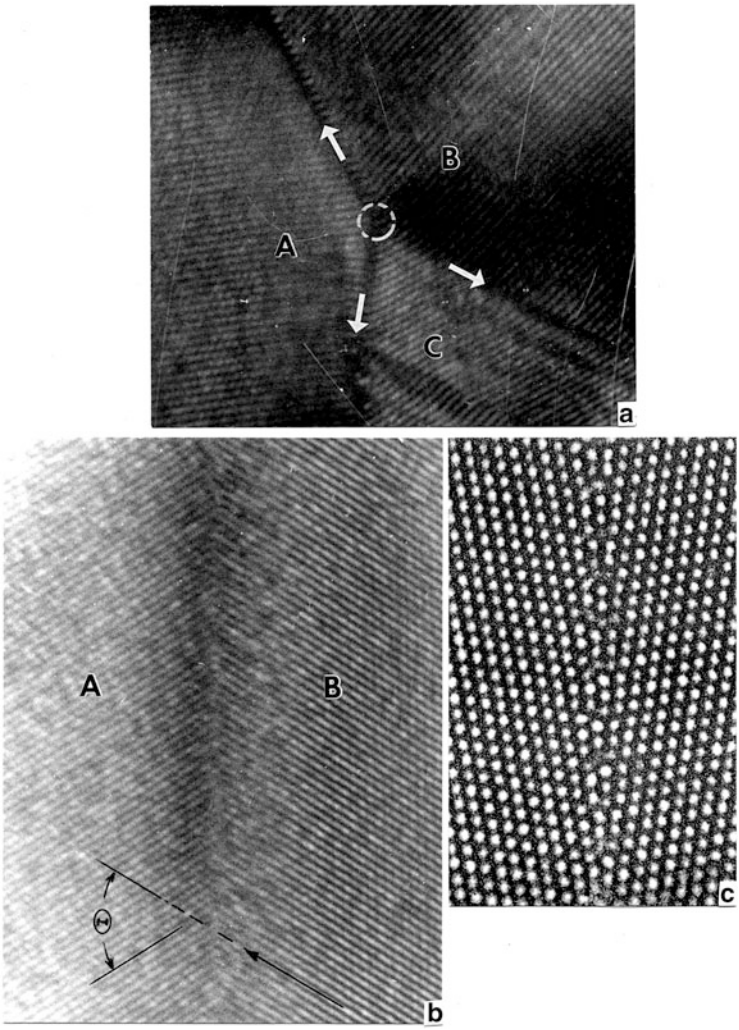


Fig. 6 High-resolution lattice and atomic images of grain boundary structure in aluminum. (**a**) Three grains forming a triple junction or triple point corresponding to Fig. 5. The *white arrows* denote interfacial free energies (grain boundary free energies) resolved in the grain boundary plane. Note that the arrow (energy) in boundary AB (γ_{AB}) "pulls" γ_{AC} and γ_{BC} symmetrically. $\gamma_{AC} \cong \gamma_{BC}$; $\gamma_{AB} > (\gamma_{AC}, \gamma_{BC})$. The misorientations are $\Theta_{AB} = 42°$, $\Theta_{AC} = 25°$, and $\Theta_{BC} = 19°$. The orientations of A = B = C = [001]. Imaged planes are (200). (**b**) [001] grains, $\Theta = 64°$. (**c**) Atomic image for (110) oriented grains. $\Theta = 18°$. (**c**) is courtesy of Ron Gronski, U.C. Berkeley

section thickness observed in the TEM is very thin and the atomic arrangements or crystal planes being resolved, especially in Fig. 6a and c, are perpendicular to the surface plane orientation (100), and not representative of thicker, more representative grain sections illustrated in Fig. 4.

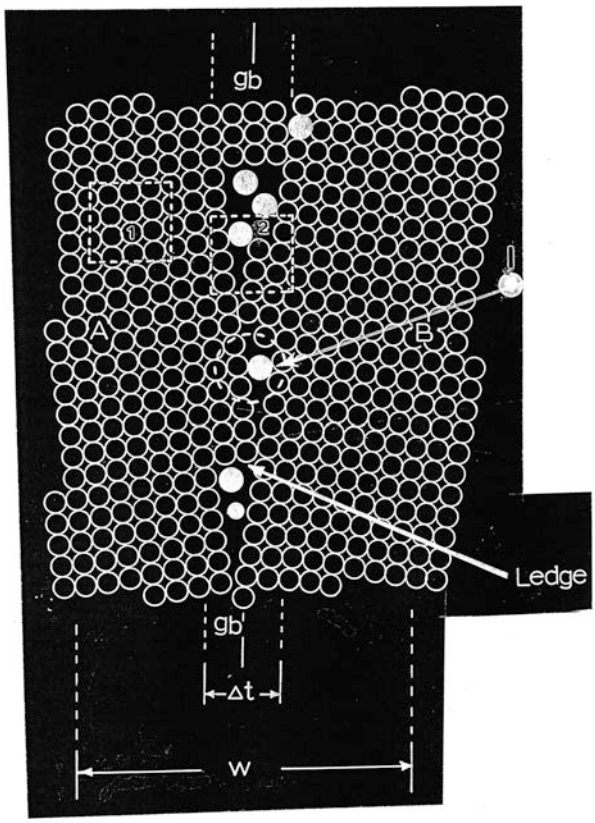


Fig. 7 Grain boundary model similar to Fig. 6c. The grain boundary plane or Gibbs dividing surface is denoted gb. Δt denotes the width of the grain boundary phase. W denotes the region which may influence electrical or other behavior. Dotted boxes in A and in the grain boundary region marked (1) and (2) denote free volume concept. I denotes impurity

Figure 7 shows a simple model representative of Fig. 6c, with essentially the same misorientation: $\Theta_{AB} \cong 18°$. The grain boundary plane, representing the Gibbs dividing surface, is shown at gb, while the grain boundary thickness is indicated by Δt, corresponding to Fig. 2. In Fig. 7, Δt is observed to be roughly 3–4 atom diameters (~1 nm). Grain boundary ledges are noted, and "I" indicates impurity atoms segregated to the grain boundary. There are two identical areas noted by dotted squares indicated by 1 in the grain (or lattice) region, and 2 in the grain boundary phase (corresponding to Δt). These regions represent free volumes or reference volumes where it can be observed, pragmatically, that the density and chemistry (or chemical potential) are different for volume (1) (or area 1) versus volume (2) (or area 2). This can have fundamental implications regarding the elastic (or shear) properties of the grain boundary phase (Δt) in contrast to the grain interior or crystal lattice as shown schematically in Fig. 4a: $E_L \neq E_{gb} \neq E_L'$; $G_L \neq G_{gb} \neq G_L'$. From first principles

$$E = E_o\left(1 - f_1 p + f_2 p^2\right), \tag{8}$$

where the elastic (or Young's) modulus, E_o, is reduced by porosity, p, and f_1 and f_2 are equal to 1.9 and 0.9, respectively. For relatively low porosity, p^2 is negligible, and

$$E/E_o = (1 - 1.9\,p). \tag{9}$$

It can also be recalled from Part VIII, Chapter "▶ Implant Materials and Structures" that for open-cellular structures having relatively larger porosities (▶ Eq. 7 of chapter "Implant Materials and Structures"),

$$E/E_o = (\rho/\rho_o)^2, \tag{10}$$

where ρ is the density of the open-cellular structure and ρ_o is the fully dense, solid structure.

Correspondingly, the porosity p, is written

$$p = (1 - \rho/\rho_o). \tag{11}$$

Sanders et al. (1997) found that the Young's modulus for nanocrystalline Cu (with a grain size of ~20 nm) decreased by ~11 % for a porosity of 0.03 (3 %). Similarly, Youngdahl et al. (1997) showed that when the density of this same copper decreased from 98.4 % to 92.5 %, the yield strength decreased from 0.84 to 0.65 GPa (a 24 % decrease).

Figure 8 shows a sketch representing a fine nanocrystalline metal having an average grain (matrix) diameter of ~1.8 nm, with a grain boundary phase dimension (Δt) of ~0.5 nm. This may represent the extreme in the nanocrystalline regime since the grain boundary width, Δt, would be roughly equal to atomic structural units in a glassy metal (Murr 1980). In Fig. 8, the grain boundary phase does not exhibit any recognizable order, and there is a significant free volume difference between the crystalline grains, and this structure essentially converges to the amorphous state (Murr 1980). Even triple junctions become nondistinct in Fig. 8 in contrast to Fig. 6. It might be noted that dislocations which could propagate in the small nanograins in Fig. 8 would simply be absorbed in the grain boundary phase. Grain boundary ledges also become essentially indistinguishable in Fig. 8 in contrast to Fig. 7.

In nanocrystalline metals and alloys, deformation is accommodated by both slip and grain boundary sliding by circumstances and mechanisms which are well documented for superplastic behavior, where grain size reduction is a critical factor. It can be observed phenomenologically that even the deformation involving shape change of a grain, or grains relative to one another, requires extension of some interfaces and displacement of the triple points connecting the grain boundaries (or interfaces) (Fig. 6). This process is facilitated by grain boundary dislocation slip and interaction, forming ledges, and the production of dislocations within the grain interiors, all preserving or conserving mass. In superplastic flow involving grain boundary sliding, temperature also plays a key role. In the simplest sense, the creep rate associated with sliding of the grain regime is expressed by

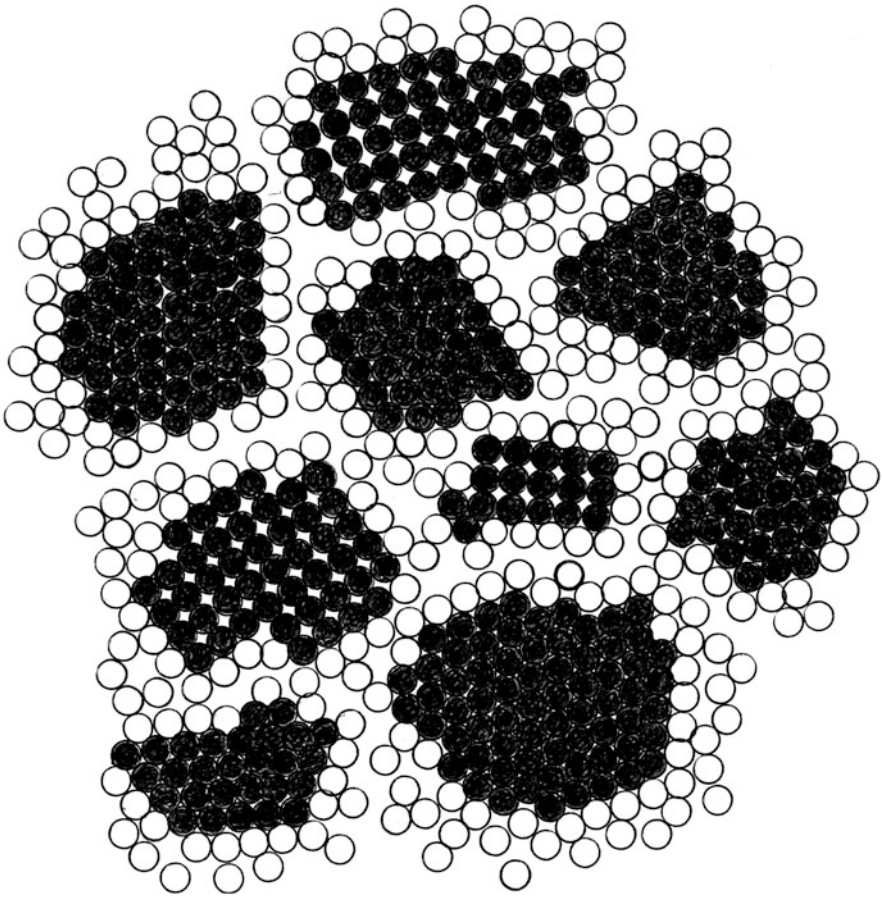


Fig. 8 Two-dimensional model illustrating the smallest nanostructure where the grain (*dark circles*) size (D) approaches the grain boundary phase (*open circles*) size (Δt). D ~ 1.5 nm. Grain (*dark*) orientations are (111) and (100)

$$\dot{\varepsilon} = \mathrm{K}\sigma\exp(-\mathrm{Q}/\mathrm{RT})/\mathrm{D}^{\mathrm{m}} \tag{12}$$

where K is a constant, σ is the applied stress, Q is the activation process, R is the gas constant, T is the temperature, D is the grain size, and m is 2 for Nabarro-Herring creep or 3 for Coble creep (Ashby 1976) (see Part VI). Assuming a somewhat phenomenological model as shown in Fig. 9, Raj and Ashby (1971) developed an expression for the grain boundary sliding rate taking into account both lattice (crystal or grain matrix) and grain boundary diffusion:

$$\dot{u} = \frac{du}{dt} = \frac{2\tau\Omega}{\pi\mathrm{kT}}\left(\frac{\lambda}{h^2}\right)\mathrm{D_L}\left(1+\frac{\pi\Delta t}{\lambda}\left(\frac{\mathrm{D_{gb}}}{\mathrm{D_L}}\right)\right), \tag{13}$$

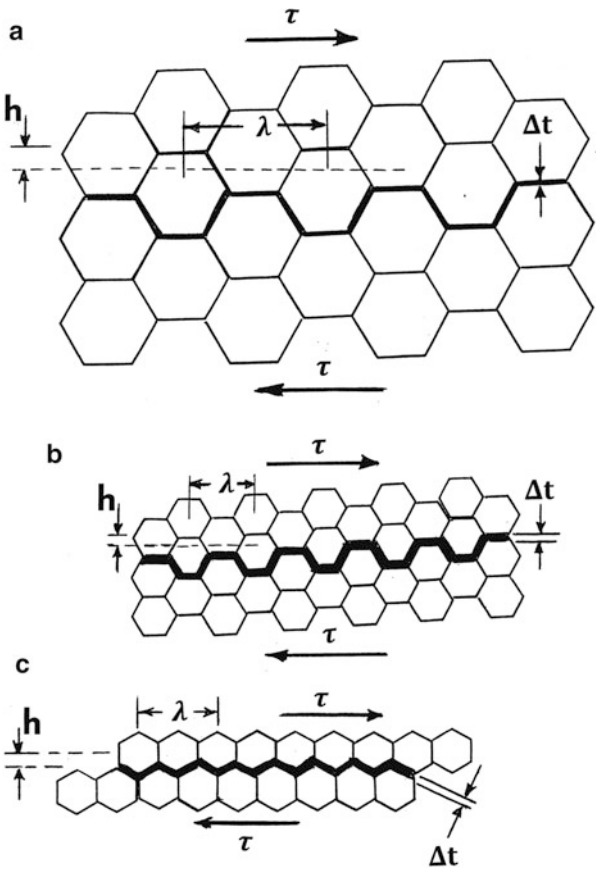


Fig. 9 Schematic representations of grain boundary sliding under shear stress, τ. Corresponding sine wave-related displacement wavelength, λ, and amplitude, h, are denoted. Grain size is denoted D and the grain boundary phase thickness is denoted Δt. (**a**) Large grain (micron) regime. (**b**) Nanograin regime. (**c**) Grain orientation is 90° from (**b**)

where τ is the shear stress; k is the Boltzmann constant; T is the temperature; Ω is the atomic volume; D_L is the crystal lattice diffusion coefficient; D_{gb} is the grain boundary diffusion coefficient; λ is the displacement wavelength and h is the amplitude factor, both of which are related to the grain size (decreasing for decreasing grain size); and Δt is the grain boundary phase thickness. λ and h can be rendered as a sine wave, and there is some small dimensional variance between the sliding geometry of Fig. 9b in contrast with Fig. 9c. If in fact each grain boundary layer generates an identical sliding rate as illustrated schematically in Fig. 9c, a shear strain rate can be defined as

$$\dot{\gamma} = \dot{u}/2h = \frac{\tau\Omega}{\pi kT}\left(\frac{\lambda}{h^3}\right)D_L\left(1 + \frac{\pi\Delta t}{\lambda}\left(\frac{D_{gb}}{D_L}\right)\right), \tag{14}$$

since in Fig. 9c, $D \cong 2\,h$, and considering that for temperatures less than half the melting temperature, $D_{gb} >> D_L$ (as intuitively implicit in Fig. 8), the term $\frac{\pi\Delta t}{\lambda}\left(\frac{D_{gb}}{D_L}\right) >> 1$ in Eq. 14;

consequently

$$\dot{\gamma} \cong \frac{8\Delta t \Omega D_{gb}}{kT}\left(\frac{1}{D^3}\right)\tau. \tag{15}$$

When $\Delta t \cong D$ is assumed for the amorphous state (Murr 1980),

$$\dot{\gamma} \cong \frac{8\Omega D_{gb} t}{kT}\left(\frac{1}{D^2}\right). \tag{16}$$

Equations 15 and 16 have the form of Eq. 12 corresponding to Coble creep and Nabarro-Herring creep, respectively.

Ideally, when $\Delta t \cong D$, the strength (or yield strength) as expressed by the Hall–Petch relation in Eq. 7 should be optimized or reach a plateau equal to the yield stress for the amorphous state. This has in fact been demonstrated by Meyers et al. (2006) who compared an empirical model considering areal fractions for the grain interior (or crystal lattice), A_L, with that for the grain boundary phase, A_{gb}, in the form

$$\sigma_y = A_L\sigma_L + A_{gb}\sigma_{gb}, \tag{17}$$

which is conceptually the same as the volume fraction representation by Margolin and Stanescu (1975) shown in Eq. 6, where $\overline{\sigma}_L = \overline{\sigma}_B$. This concept can be qualitatively applied to the representation of the grain matrix and the grain boundary phase in Fig. 2. Here the areal fractions are expressed by

$$A_{gb} = \frac{D^2 - (D - \Delta t)^2}{D^2} = 2\left(\frac{\Delta t}{D}\right) - 2\left(\frac{\Delta t}{D}\right)^2, \text{and} \tag{18}$$

$$A_L = 1 - A_{gb} = 1 - 2\left(\frac{\Delta t}{D}\right) - 2\left(\frac{\Delta t}{D}\right)^2, \tag{19}$$

From Eq. 17, the yield stress becomes

$$\sigma_y = \sigma_L + 2\left(\sigma_{gb} - \sigma_L\right)\left(\frac{\Delta t}{D}\right) - \left(\sigma_{gb} - \sigma_L\right)\left(\frac{\Delta t}{D}\right)^2 \tag{20}$$

If it is assumed that $\Delta t = K\sqrt{D}$, then Eq. 20 takes the form

$$\sigma_y = \sigma_L + 2\Delta\sigma/\sqrt{D} - \Delta\sigma/D \tag{21}$$

which is similar to the Meyers et al. (2006) expression that also assumed that $\Delta t = K\sqrt{D}$, where K is a constant, and $\Delta\sigma = (\sigma_{gb} - \sigma_L)$. However, their derivation was based on dislocation hardening at the grain boundary in micron-size polycrystal grains as shown in Fig. 4b assuming a constant

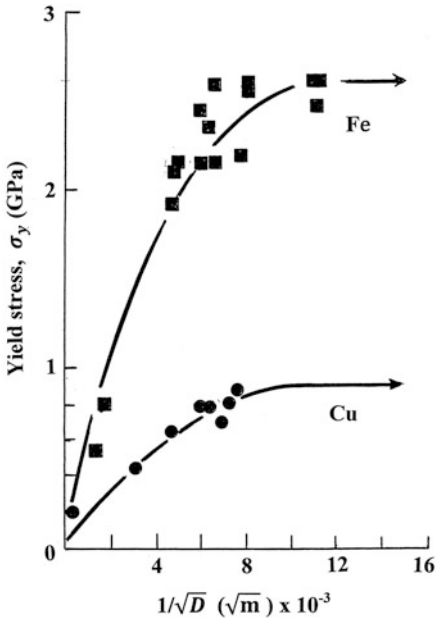


Fig. 10 Plots of σ_y versus $1/\sqrt{D}$ for (a) Fe and (b) Cu with fitted line modeled by Meyers et al. (2006) in the form of Eq. 21 (Adapted from Meyers et al. (2006))

dislocation-hardened zone having a thickness t, which in the present situation relating to the grain boundary phase represented in Fig. 2 equals $\Delta t/2$. However, as illustrated in the experimental data in Fig. 3, the dislocation density near the grain boundary increases with increasing strain, but at a different rate than the grain interior. Nonetheless, Meyers et al. (2006) compared this model equation (of the form of Eq. 21) to experimental data for nanocrystalline Cu and Fe as illustrated in Fig. 10. The plateaus shown are consistent with the measurements of microindentation hardness for amorphous Cu and Fe of ~3 GPa for Cu and ~9 GPa for Fe, where, from Eq. 7, $\sigma_y \cong Hv/3$.

Despite the varied schematic and graphical perspectives of nanocrystalline materials provided in Figs. 2, 3, 4, 5, 6, 7, 8, 9, and 10, it is not possible to provide a fundamental understanding of deformation dynamics in nanocrystalline metals, particularly at the very smallest dimensions as implicit in Fig. 8. In addition, there have been far more applications of nanocomposites such as the Wootz steel precursors and performance of fabricated swords described in Chapter "► Serendipitous Nanotechnology in Antiquity" than contemporary applications of nanocrystalline metals such as Cu or Fe whose strength performance is illustrated in Fig. 10. Computer simulations have provided some indications of the dynamic response of nanocrystalline metals and are summarized in Meyers et al. (2006).

Before reviewing some of the conclusions reached by examining computer simulations, it might be useful to explore a few contradictions observed for nanocrystalline metals, especially the role of dislocations and dislocation-related

mechanisms and microstructures. It might be recalled that in the micron range of (grain size) polycrystalline metals, deformation, such as the variations of shock-wave deformation of Cu as described in Part V (Chapter "▶ Strengthening by Crystal Imperfections," Fig. 9), the formation of stacking faults, twins, and microbands decreases for decreasing grain size. In addition, Al, with a stacking-fault free energy more than twice that for Cu (Murr 1975), does not exhibit stacking faults or twins for any mode of deformation in the micro-grain range. However, Chen et al. (2003) illustrated twins in nanocrystalline Al, while molecular dynamics (MD) simulations for Al have subsequently shown profuse partial dislocation emission from the grain boundary phase forming stacking faults and twins in ~10 nm grain size nanocrystalline Al (Froseth et al. 2004; Derlet et al. 2003, 2004). The implications are that at some threshold nanograin size, perfect dislocations decompose into partial dislocations, and there is a suppression of stacking-fault free energy and corresponding twin boundary free energy.

Profuse twinning has been shown to occur in the electroplating growth of Cu. In addition, the development of nanotwin structures in nanograin Cu produces a significant elevation of the microindentation hardness and the corresponding yield stress as implicit in Eq. 7. This phenomenon is illustrated in the work of Lu et al. (2004) and You et al. (2011) reproduced in Fig. 11a and b, respectively. Figure 12 illustrates the coherent {111} nanotwins in roughly equiaxed nanograin Cu with a grain size averaging ~400 nm. It can be observed in Fig. 11a that 15 nm coherent twin spacing (as observed in Fig. 12) increases the strength by more than twice that for nanograin Cu (D 100 nm). In addition, the elongation (ductility) for the nanotwin Cu in Fig. 11a is more than a factor 10 larger than nanocrystalline (nanograin) Cu. This indicates that coherent twin boundaries promote a different deformation mechanism than nanograin boundaries. In addition, Lu et al. (2004) showed that the nanotwinned Cu had an electrical resistivity only slightly higher than microcrystalline Cu, while nanograin (D ≥100 nm) Cu had an electrical resistivity more than three orders of magnitude higher.

On comparing coherent {111} nanotwins in equiaxed, nanocrystalline Cu (with a grain size of ~400 nm) with {111} nanotwins in columnar-grained Cu with a length of 30 μm and an aspect ratio of 10 (average width ~300 nm) and ~48 nm twin width, You et al. (2011) assumed a Hall–Petch (H–P) model for 15 nm twins in 400 nm equiaxed grains as in Fig. 11a by replacing the grain size, D, with twin thickness, λ:

$$\sigma_y \sigma_o + K_{tb}/\sqrt{\lambda} \quad \underline{(H-P)} \tag{22}$$

where K_{tb} is the twin-related material constant. For columnar-grain Cu containing {111} nanotwins, they assumed a so-called confined-layer slip (CLS) model outlined by Misra et al. (2005):

$$\sigma_y = \sigma_o + \beta \frac{G_b}{\lambda} ln\left(\frac{\alpha\lambda}{b}\right) \quad \underline{(CLS)}, \tag{23}$$

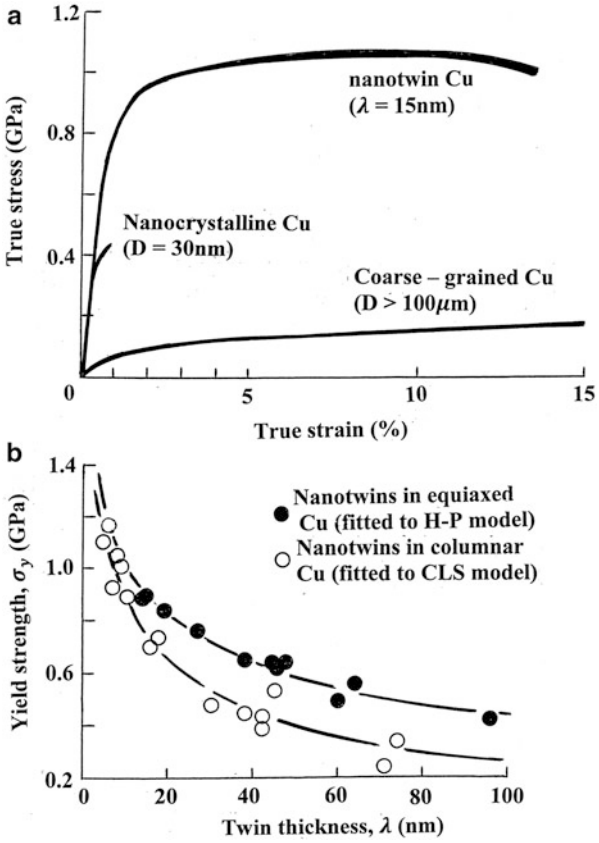


Fig. 11 (**a**) True stress–strain curves for electrodeposited Cu with nanotwin structure, nanocrystalline Cu, and coarse-grain, polycrystalline Cu with grain or twin dimensions indicated (Adapted from Lu et al. (2004)). (**b**) Yield strength variation ($H_v/3$) of columnar, nanotwinned Cu compared with equiaxed, nanotwinned Cu along with fitted models: Hall–Petch (H–P) and confined-layer slip (CLS) (Adapted from You et al. (2011))

where α and β are materials constants, G is the shear modulus, b is the scalar value of the Burgers vector, and λ is the twin thickness. These models were fit to the data reproduced in Fig. 11b which is discussed in detail by You et al. (2011). Figure 11b indicates that for twin thicknesses approaching the nanograin sizes in the schematic views of Fig. 8, the yield strength would converge to a value of $\geq$ 1.5 GPa.

Taken together, the experimental and computational evidence relating to the deformation behavior of nanocrystalline metals point to complex interplay between the behavior of dislocations in the grain interior (crystal lattice) and the interfacial phase. In this sense, ledges become indistinguishable at the small nanograin size and grain boundary sliding assumes varying roles, intermixed with crystalline slip. Dislocations do not pile up at "boundaries" in nanocrystalline materials, and cross-slip is not observed in the crystal lattice. It may also be that impurities or impurity atoms play a much more important role in the grain boundary phase of nanocrystalline metals than in micron-range polycrystalline metals. This is illustrated in Fig. 13 which is a modified cartoon of Fig. 8. Figure 13 may also represent the prospects for stabilizing the nanocrystalline regime as described in recent work of Weertman (2012).

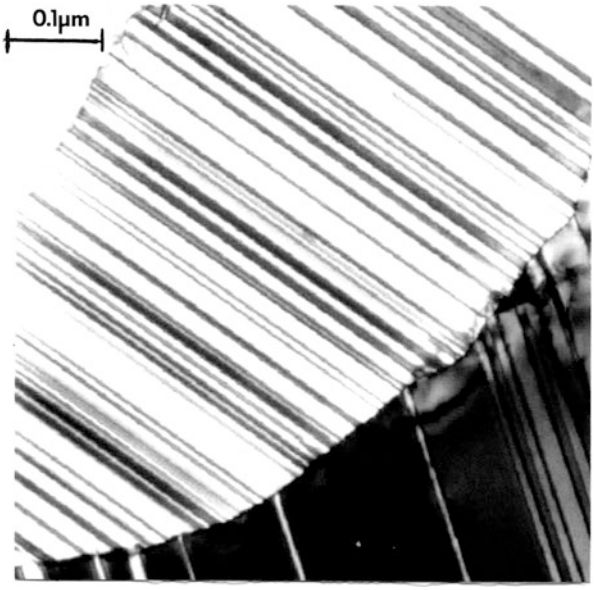


Fig. 12 TEM image showing coherent {111} nanotwins in electrodeposited Cu (Courtesy of Dr. Lei Lu, Shenyang National Laboratory for Materials Science (Adapted from Lu et al. (2004)))

It is also interesting to consider the implications of the very significant difference in strength implicit in Fig. 11a for elastic (brittle) regimes such as ceramics in contrast to metals and alloys. The imposition of some degree of grain boundary sliding in polycrystalline, and especially nanocrystalline ceramics, can improve the ductility, if even by a few percent elongation. In fact, varying degrees of superplastic behavior have been demonstrated in oxide ceramics and the like (see, e.g., Mayo (1997)).

In retrospect, nanocrystalline materials deformation might be viewed in simple cartoons as shown in Fig. 14. In Fig. 14(1) deformation is dominated by dislocation source activation and slip in the crystal lattice. In Fig. 14(2), the grain boundary phase dominates: grain boundary dislocations facilitate grain boundary slip (or sliding). In Fig. 14(3), some combination of (1) and (2) are involved. The role that these processes have on extreme deformation processes in particular will be discussed in more detail in Part XIII (Chapter "▶ Ballistic and Hypervelocity Impact and Penetration") which follows.

Mechanical Properties Influenced by Regime Size for Nanostructured Materials

In nearly every structural regime, size matters. It has long been known that even for engineering materials or polycrystalline metals and alloys in particular, it is necessary to assure that for tensile testing specimens, the specimen thickness (Δ) (or dimensions) should be at least 7 grain diameters (D): $\Delta/D \geq 7$. However, Hansen (1977) observed a size effect in Al polycrystalline specimens

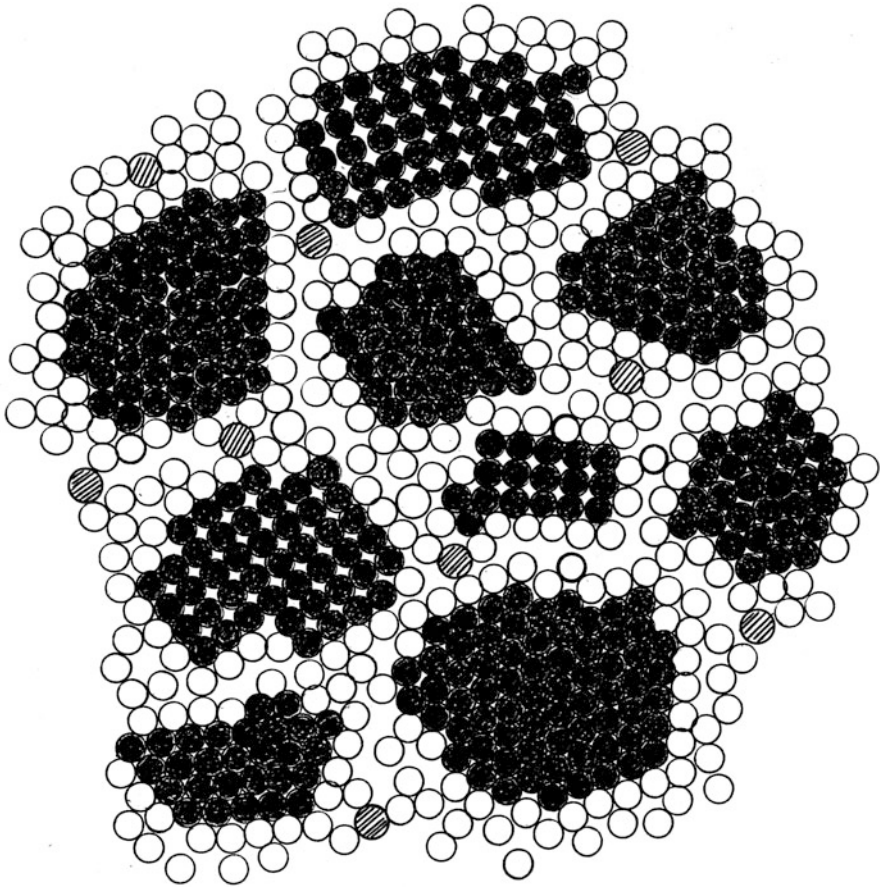


Fig. 13 Nanocrystalline metal model for very small nanograins as shown in Fig. 8 with grain boundary phase additions (impurities) which can alter the phase stability, chemical potential (and corresponding interfacial free energy), and diffusivity

with $\Delta/D < 15$. In addition, it has been known that deformation will differ if the specimen size for a single crystal becomes very small (whisker size). In other words, a combination of internal and external length scales contributes to the mechanical properties. For example, as illustrated for the tensile straining of macroscopic single-crystal cylinders in ▶ Fig. 9 of chapter "Line Defects: Dislocations in Crystalline Materials," dislocation sources in the surface predominate, and dislocation density, especially at lower strains, will be higher near the surface sources in the same way grain boundaries are hardened by source production as illustrated schematically in Fig. 4b and the experimental measurements for 304 stainless steel reproduced in Fig. 3. Even in polycrystalline metals and alloys having a uniform cross section, such as a wire, the deformation will change as the grain boundaries are annihilated or restructured to lower the total interfacial free

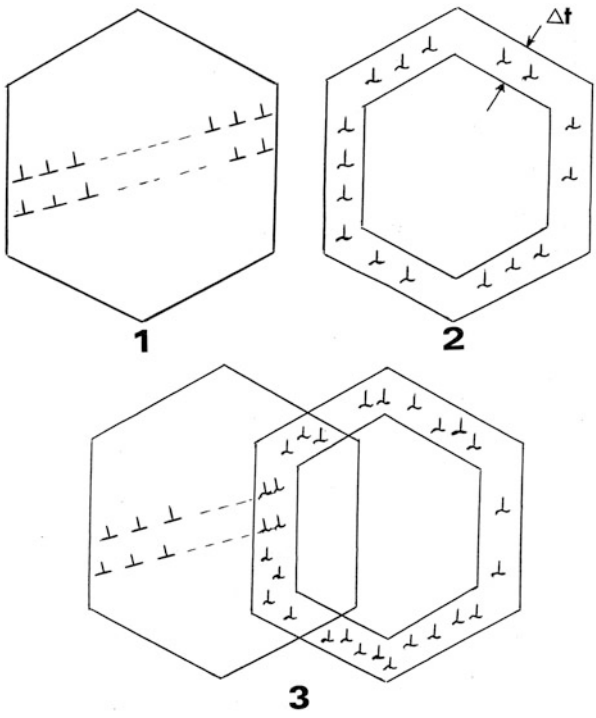


Fig. 14 Lattice and grain boundary dislocation models. (**1**) Lattice dislocation mechanisms: source production, slip, pileup. (**2**) Grain boundary dislocation slip and interaction. (**3**) Combined lattice slip–grain boundary phase slip/sliding. Note that any of these grains can also be deformed or distorted

energy. This occurs for wires forming connected single-crystal grains having grain boundaries traversing the section normal to the wire axis as shown for so-called bamboo structure in Chapter "► Planar Defects: Crystal Interfaces," Fig. 5. Such structures allow simple shear along the normal grain boundaries, and in the case of filaments in light bulbs, this shear produces high resistance regions on the grain boundary and creates hot spots which cause filament failure as noted previously. Like nanocrystal structures, it is important to stabilize the grain boundaries and prevent energy minimization. These features are illustrated in the comparative schematics in Fig. 15, and the SEM images shown in Fig. 16. It can be observed that Fig. 15c can also illustrate Fig. 16a at the specimen edge and therefore represent polycrystal grain boundaries traversing the sample thickness, as well as the perpendicular boundaries shown in the bamboo structure wire in Fig. 16b. In addition, Fig. 15c can also represent grain boundary triple junctions traversing the specimen thickness. These structures can function differently from the grain boundary plane, and as the grain size is reduced forming boundaries perpendicular to the specimen surfaces as represented in Fig. 15c, the density (or number) of triple junctions intersecting unit area of surface would also increase. This would affect diffusion as well as plastic deformation (or plastic flow) as discussed by Fedorov et al. (2002). Triple junctions, as illustrated in the cartoon of Figs. 8 and 13, can also act as special nucleation points for dislocations, and they affect grain boundary sliding as well.

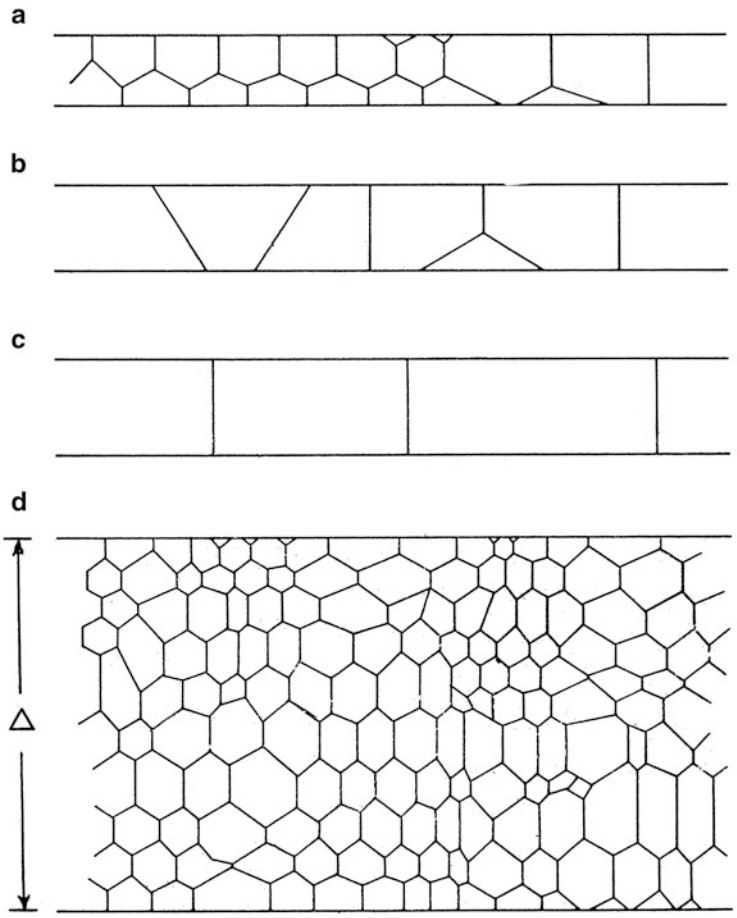


Fig. 15 Grain-in-thickness schematics. (**a**–**d**) illustrate several examples of simple specimen thickness views with grain and grain boundary configurations. (**a**–**c**) illustrate energy reduction by grain (interface) area reduction. (**d**) illustrates a section thickness (Δ) containing ~8 grain diameters (D): $\Delta/D \sim 8$

Greer et al. (2012) have described deformation mechanisms in nanostructured materials in the context of constituent architectures as this applies to cellular architectures where various length scales apply to components of the architecture, such as cross sections, wall thicknesses, etc., as well as their microstructural and nanostructural compositions. This included compression testing of 50 μm diameter Ni pillars in contrast to hollow pillars of the same dimension, with a wall thickness of 150 nm. It was also noted that for grain sizes ~10 nm in 50 nm diameter Ni pillars, strength was similar to nanocrystalline bulk Ni having the same grain size. It was mentioned that the strength of nanocrystalline platinum also did not deviate significantly from bulk (yield strength) down to $\Delta/D = 8$.

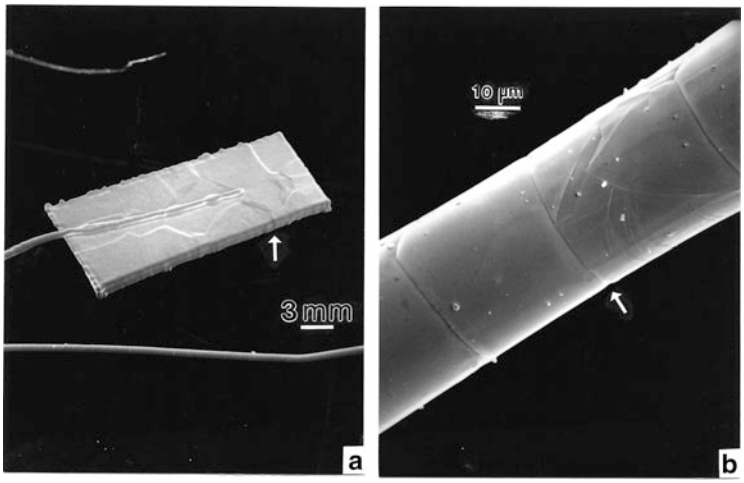


Fig. 16 SEM images showing grain boundaries traversing specimen thickness. (**a**) Rectangular 304 stainless steel specimen containing grains with grain boundaries traversing the sample thickness (*arrow*). (**b**) Grain boundaries perpendicular to wire axis for 304 stainless steel forming bamboo structure

In a more recent study of the influence of sample thickness on the tensile properties of pure Cu with different grain sizes, Yang and Lu (2013) found that for coarse-grained Cu having a grain size of ~82 μm, $\Delta/D = 6$ implied consistent measurements, while for fine-grained Cu having a grain size of ~13 μm, $\Delta/D = 18$ was observed to be the critical value. Their conclusion was that dislocation activities are related to specimen thickness, which is implicit in Figs. 15 and 16, as well as Fig. 14. Regardless of the length scales involved, it appears prudent to assure Δ/D 7 as a lower limit, or critical value, even for (and perhaps especially for) nanocrystalline materials, particularly nanocrystalline metals and alloys.

References

Ashby MF (1970) The deformation of plastically non-homogeneous materials. Phil Mag 21:399–424

Ashby MF (1976) Boundary defeats and atomistic aspects of boundary sliding and diffusional creep. Surf Sci 31:498–543

Ashby MF, Greer A (2006) Metallic glasses as structural materials. Scripta Metall 54:321–326

Chen MW, Ma E, Hemker KJ, Sheng H, Wang Y, Cheng X (2003) Deformation twinning in nanocrystalline aluminum. Science 300:1275–1277

Derlet PM, Hasnaoui A, Van Swygenhoven H (2003) Atomistic simulations as guidance to experiments. Scripta Mater 49:629–635

Derlet PM, Froseth AG, Van Swygenhoven H (2004) The influence of twins on the mechanical properties of nc-Al. Acta Mater 52(20):2259–2268

Fedorov AA, Gutkin MY, Ovid'Ko IA (2002) Triple junction diffusion and plastic flow in fine-grained materials. Scripta Mater 47:51–55
Froseth AG, Derlet PM, Van Swygenhoven H (2004) Dislocations emitted from nanocrystalline grain boundaries: nucleation and splitting distance. Acta Mater 52(20):5867–5870
Gleiter H (1981) Materials with ultrafine grain size. In: Hansen N (ed) Deformation of polycrystals: mechanisms and microstructures. Riso National Laboratory, Roskilde, p 15
Greer JR, Jang D, Gu XW (2012) Exploring deformation mechanisms in nanostructured materials. JOM 64(10):1241–1252
Hansen N (1977) The effect of grain size and strain on the tensile flow stress of aluminum at room temperature. Acta Metall 25(8):863–869
Langford G (1977) Deformation of pearlite. Metall Trans A 8A:861–875
Lashmore DS, Jesser WA, Schladitz DM, Schladitz HJ, Wilsdorf HGF (1977) Microstructural investigation of polycrystalline iron whiskers. J Appl Phys 48(2):478–480
Li JCM (1976) Defect mechanisms in the deformation of amorphous materials, Chap. 15. In: Murr LE, Stein C (eds) Frontiers in materials science. Marcel Dekker, New York, pp 527–548
Lu L, Shen Y, Chen X, Qian L, Lu K (2004) Ultrahigh strength and high electrical conductivity in copper. Science 304:422–426
Margolin H, Stanescu MS (1975) Polycrystalline strengthening. Acta Metall 23: 1411–1420
Mayo MJ (1997) High and low temperature superplasticity in nanocrystalline materials. Nanostruct Mater 9:717–726
Meyers MA, Mishra A, Benson DJ (2006) Mechanical properties of nanostructured materials. Prog Mater Sci 51:427–556
Misra A, Hirth JP, Hoagland RG (2005) Length-scale-dependent deformation mechanisms in incoherent metallic multilayered composites. Acta Mater 53:4817–4824
Murr LE (1975) Interfacial phenomena in metals and alloys. Addison-Wesley, Reading
Murr LE (1980) A form of Hall–Petch law for the strength of glassy metals. Phys Stat Sol (a) 61: K101–K107
Murr LE, Wang S-H (1982) Comparison of microstructural evolution associated with the stress–strain diagrams for nickel and 304 stainless steel: Ar electron microscope study of microyielding and plastic flow. Res Mech 4:237–274
Rack HJ, Cohen M (1976) Influence of recovery on the tensile behavior of highly-strained iron alloys. Chap. 10. In: Murr LE, Stein C (eds) Frontiers in materials science. Marcel Dekker, New York, pp 365–389
Raj R, Ashby MF (1971) Grain boundary sliding and diffusional creep. Metall Trans 2A:1113–1127
Sanders PG, Eastman JA, Weertman JR (1997) Elastic and tensile behavior of nanocrystalline copper and palladium. Acta Mater 45:4019–4025
Thompson AW, Flanagan WF (1973) The dependence of polycrystal work-hardening on grain size. Acta Metall 21:1017–1028
Weertman JR (2012) Retaining the nano in nanocrystalline alloys. Science 337:921–922
Wilsdorf HGF, Inall OT, Murr LE (1978) Microstructural observations in high strength polycrystalline iron whiskers. Z Metallkunde 69(11):701–705
Yang L, Lu L (2013) The influence of sample thickness on the tensile properties of pure Cu with different grain sizes. Scripta Mater 69:242–245
You ZS, Lu L, Lu K (2011) Tensile behavior of columnar grained Cu with preferentially oriented nanoscale twins. Acta Mater 59:6927–6937
Youngdahl CJ, Sanders PG, Eastman JA, Weertman JR (1997) Compression yield strengths of nanocrystalline Cu and Pd. Scripta Mater 37:809–881

Performance, Applications, and Health Concerns of Nanomaterials

Contents

Abstract

After more than 2,000 years, chrysotile asbestos continues to pose a health risk in spite of its roughly 3,000 or more applications. Contemporary nanomaterials, especially nanoparticulate materials, are produced in millions of metric tons to create products ranging from cosmetics to composites. While many health concerns for contemporary nanomaterials are historically grounded in documented episodes involving chronic respiratory and digestive health issues, including cancers, there is limited associations with many nanoparticulate materials, although cytotoxic assays performed on a variety of nanoparticles and nanoparticle aggregates using human respiratory or lung cell models almost invariably demonstrate some degree of proinflammatory response, and more extensive concerns.

Introduction

There have been many examples of nanomaterial applications, both historical and contemporary, discussed throughout this handbook. These have included Egyptian, Chinese, and Mayan pigments and dye-supporting nanoparticles, inks such as black carbon, laminar composites in the fabrication of damask and similar pattern swords, and asbestos. These applications, as a consequence of their performance and multifunctionality, continue today. The current global production of asbestos continues (in 2013) to be ~2 million metric tons, with Russia and

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_48

China being the leading producers, although many products continue to be produced and marketed in the USA. Asbestos–cement (transite) products, fire protection products, insulation products, and filters (including water and food products) dominate the thousands of asbestos-related products, many of which are composites. Black carbon (BC) production globally amounts to ~9 million metric tons, the bulk of which is used in tire (▶ Fig. 9 in chapter "Examples of Man-Made Composite Structures") and related rubber production as a strengthening agent, while another portion of BC production is dedicated to black ink manufacture.

Other prominent nanoparticle applications include ~8 million metric tons of rutile (TiO_2) used in paint, plastic, and paper products as well as lotions, especially sunscreens. Talc finds similar applications consuming ~0.6 million metric tons. Fe_2O_3 (red hematite) in the particle size range of 200–500 nm finds extensive use in red-based cosmetics and pigment markets which consume roughly one million metric tons. Alumina (Al_2O_3) nanoparticles as well as nanofibers find extensive use in pigment, paint, and fuel additive products as well as polymer-reinforced composites and hot-pressed ceramic composites such as Al_2O_3/SiC. These products, in the aggregate, consume ~30 million metric tons of alumina nanoparticulates, including abrasive powder applications. Considerable quantities of noble metal catalyst nanoparticles are found in a range of such commercial products such as automobile catalytic converters and the like.

It is estimated that global production of carbon nanotubes (SWCNTs and MWCNTs) reached roughly 7,000 metric tons in 2012, with 100 companies involved in production or product development. DeVolder et al. (2013) recently summarized carbon nanotube commercial applications which included composite products, especially polymer and epoxy resin matrix composites, coatings and films, microelectronics; energy systems, including storage; and biotechnology, including biosensors. Carbon nanotube–polymer composites include a wide range of commercial products (Grady 2011), and the ability to grow fibers in centimeter lengths and create yarn products (Tran et al. 2011) has enhanced their application in growing numbers of commercial area. Carbon nanotube strengths have been measured as high as 100 GPa, roughly ten times than many other commercial (industrial) fiber materials. Carbon nanotubes composed roughly 28 % of the market share of overall nanomaterial demand in 2012. Even graphene has begun to find applications in numerous composite materials systems as recently discussed by Koratkar (2013).

An emerging area of colloidal suspensions involves nanofluids: suspensions of 1–100 nm nanoparticles in water, oils, or glycols. Applications include novel optical properties, dielectric liquids, conduction and convection mass transport media, waste heat collections and solar energy harvesting, and extensive use in the medical field (Taylor et al. 2013). These can include quantum dots dispersed in a fluid used in medical imaging and suspensions of gold nanoparticles (1–10 nm diameter) in solution with body fluids which target key areas in cancer treatment protocols, including hyperthermia treatments which selectively target cancer cells (see also Aguilar 2012).

Table 1 Nanomaterial product type, application examples, and markets[a]

Type	Application examples	Estimated products (2010)
Nanoparticles	Catalysts, sunscreens (ZnO, CeO_2, TiO_2), drug carriers, Ag antibacterial wound dressings, colloidal Ag antibacterial–antiviral–antifungal, fungicides, pigments, cosmetics (Fe_2O_3), inks, paints, MnO_2 battery material, etc.	300
Nanotubes (including carbon nanotubes)	Composites, filtration systems, cosmetics, protective clothing, conductive polymers, electrodes, biomedical scaffolds, catalytic membranes, electromagnetic shielding, sports products, etc.	>250[b]
Nanoporous materials	Catalysts, catalysis, thermal insulation, electrode materials, environmental filters, membranes, controlled release drug carriers (atomizers), etc.	40
Fullerenes	Lubricants, catalysts, composite reinforcements, medical applications, etc.	30
Quantum dots	Semiconductor products, photosensors, telecommunication, light-emitting displays, etc.	35
Nanostructured materials	Composite systems, sporting goods, coatings (various), membranes and filtration systems, photochromic glass and related systems, etc.	30
Nanofibers	Castings and joining, optical systems, filtration, membranes, reinforced composites, protective clothing, etc.	25
Nanocapsules	Drug delivery, food industry, agrochemicals, cosmetics, adhesive components, textiles (as aromatic additives to leather or other textiles)	25
Nanowires	Light-emitting diodes, chemical separation, data storage, metallic interconnects/nano- and quantum devices	20

Primary nanomaterial market sectors	Estimated market focus (2010)
Medical/pharmaceutical	30 %
Chemicals and advanced materials	29 %
Information technology	21 %
Energy	10 %
Automotive	5 %
Aerospace	2 %
Textiles	2 %
Agriculture	1 %

[a]Based on updated data in Murr et al. (2008)
[b]Includes only a few asbestos (chrysotile) applications. There are many hundreds worldwide especially in asbestos–cement composites and fire retardant systems

Table 1 summarizes some of the more contemporary nanomaterial (primarily nanoparticle precursor) applications or products. Additional nanoparticle applications will be discussed later in Part XV which deals with optical, photonic, and related magnetic properties and performance issues, respectively.

Applications and Performance of Bulk Nanomaterials

As with most other materials, nanomaterials or more specifically nanocrystalline materials applications tend to be limited in some cases by performance. As might be expected, performance is also limited by processing limitations. However, optimization of processes to create nanocrystalline materials (chapter "► Synthesis and Processing of Nanomaterials") has advanced performance, and applications of bulk nanocrystalline materials, especially metals and alloys, have proliferated in recent years (Koch et al. 2005; Whang 2011). Unlike nanoparticles of various forms which are manufactured in many thousands of metric tons (Table 1), bulk nanocrystalline materials and processes to fabricate them are measured in selected applications in the automotive, aerospace, and specialty industries involving specialized or novel commodities or modifications of existing commodities, such as friction-stir processing of existing (commercial) plate products and the like. Other forms of surface coatings include spray coatings having nanograin structures, as well as electrodeposited layers which can have hardnesses of more than five times than the more conventional (or micron) grain size material and up to 200 times better wear resistance (Zehetbauer and Zhu 2009) (see also ► Fig. 11 in chapter "Mechanical and Other Properties of Nanocrystalline Materials"). In addition, billet production using Equal Channel Angle Processing (ECAP), mechanical alloying, accumulative roll bonding (the continuous rolling of metal or alloy laminates), and other forms of severe plastic deformation (SPD) is also contributing to the production of a variety of nanocrystalline metals and alloys in particular having a range of applications. These include nanostructured steels for automotive body structures, which are functionally related to damask and samurai sword fabrication more than 1,000 years ago. Indeed, Reza et al. (2013) have recently described a nanostructured aluminum/copper composite fabricated by accumulative roll bonding and folding, which emulates exactly the damask swordsmithing process (chapter "► Classifications and Structures of Nanomaterials").

It might be mentioned that in this process of repeated folding and roll bonding of different laminates, as well as other SPD processes (chapter "► Synthesis and Processing of Nanomaterials"), repeated, extreme deformation creates nanograins having high dislocation densities as well as heavily deformed grain boundaries or grain boundary phases which are often referred to as nonequilibrium grain boundaries. These, nanostructures, as recently discussed by Weertman (2012), are often metastable and prone to some recrystallization and grain growth even at room temperature as a consequence of their high energy and associated high entropy. However, such phenomena are often used in other functional regimes such as the storage of hydrogen in high-energy ball milling of nanocrystalline metal hydrides (Koch 1997).

The role of SPD in all forms of extreme deformation such as machining, explosive and friction welding, ballistic and hypervelocity impact and penetration, etc. and the formation of shear localization in adiabatic shear bands also involve these high-energy regimes. These will be explored in more detail in Part XIII which follows.

Health and Safety Concerns

While it might be stated at the outset that the health effects of nanomaterials, particularly nanoparticulates and nanoparticle aggregates, are to some extent circumstantial and anecdotal, the processing and applications of nanoparticulates in particular should be conducted with concerns for the environment and human health. The European Commission's Scientific Committee on Emerging and Newly Identified Health Risks (especially those associated with nanotechnology developments) noted in 2012 their principal recommendation: "safety by design."

History, in regard to nanoparticulates, has already set a sobering example to health effects of asbestos (especially chrysotile, asbestos nanotubes), black carbon/soots, NiO (dust), and a host of other industrial nanoparticulates and their products. Asbestos is of course the most notable in the context of nanoparticulate-related, multifunctional products, since most of the world continues to deal with the varieties of cancers caused by asbestos, in addition to the related chronic respiratory problems. It is remarkable that even though asbestos as a recognizable, chronic health issue was described nearly 2,000 years ago by Strabo and the Greek historian Pliny the Elder (Alleman and Mossman 1997), it was not until the 1930s in England when industry regulations were first enacted in response to high incidences of fibrotic lung damage amongst British asbestos milling and other manufacturing sector workers.

Today, the paramount concern for asbestos is its release from composite and other products. In a broad sense, release is an important issue for many other more contemporary nanoparticulates such as black carbon (BC) in tires which wear progressively and release the BC and TiO_2 in paints on many outdoor structures which erode and release the TiO_2 (as illustrated in ▶ Fig. 2 in chapter "Classifications and Structures of Nanomaterials"), as well as its release from sunscreens and other lotions, often into the groundwater, along with ZnO and CeO_2, both of which are similarly released, along with many similar nanoparticulates. While little is known regarding these released nanoparticulates which may enter the food chain, a great deal is known regarding their respiratory effects in ambient and more sequestered air environments. Moreover, accumulations of numerous, released nanoparticulates may pose masking effects similar to chrysotile asbestos and talc, which can occur in the groundwater from natural geological deposits or in contaminated talc products. High incidences of stomach cancer in Asian and Chinese locations over the past decades were often blamed on talc which is used to prevent clumping of hydrated (and correspondingly hygroscopic) rice, and when not properly washed from the rice, can be ingested along with the rise. However, studies showed that in most cases of such stomach cancers, that talc was contaminated with chrysotile asbestos (Ward 2005). Similar asbestos contamination has been associated with skin irritations connected with various, compromised talc-powder products.

While the production quantities of carbon nanotubes is at least a thousand times less than many other nanoparticulates, including black carbon, its fibrous (nanotube) structure in some respects emulates that of chrysotile asbestos nanotubes as

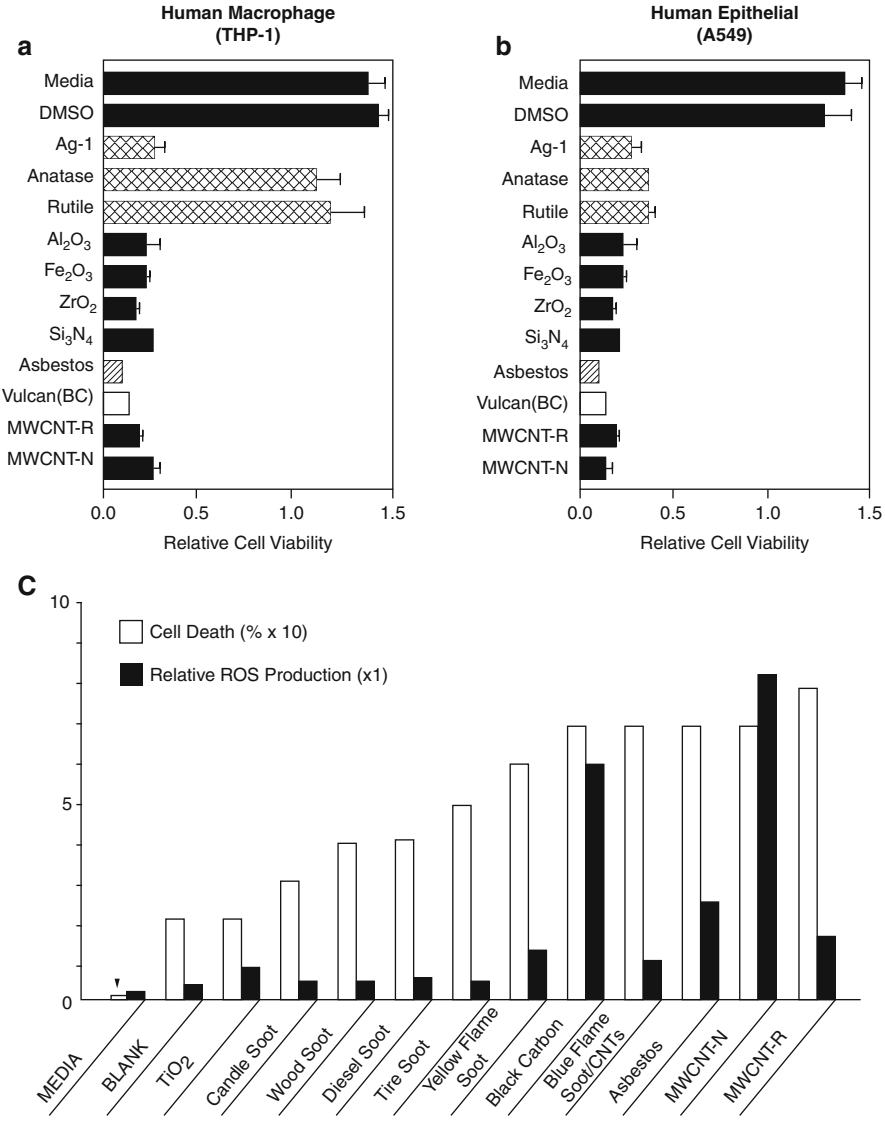


Fig. 1 (**a**) and (**b**) show comparative cytotoxicities for manufactured nanoparticulate materials for human macrophage and epithelial cell assays. Media and DMSO are controls. After Soto et al. (2007). (**c**) Comparison of cell death (in percent) and relative ROS production for nanoparticulate soots and other nanoparticulates in 48 h human epithelial (A549) cell cultures. After Garza et al. (2008)

implicit on comparing ▶ Figs. 2 and ▶ 3 (chapter "Examples of Materials Science and Engineering in Antiquity"), as earlier described by Service (1998). However, carbon nanotubes, especially multiwall carbon nanotubes, are somewhat ubiquitous in the environment, especially indoor environments where the combustion of

natural gas or propane can produce accumulations of both fullerenes and MWCNTs (see ▶ Fig. 2 of chapter "Synthesis and Processing of Nanomaterials"). Soots are also ubiquitous but differ from commercial BC in their adsorbed polycyclic aromatic hydrocarbon fragments, which at high molecular weights (MW >200) are carcinogenic and mutagenic (Murr 2012).

Although there have been only a few in vivo studies involving nanoparticulates, including carbon nanotubes, there are numerous in vitro assay studies (Nel et al. 2006; Pope and Dockery 2006; Murr 2012; Murr et al. 2008). The advantage of such studies, in contrast to in vivo studies, lies in the fact that although human cell types, and especially respiratory lung-response cells, can be utilized, specific cell types eliminate the variance often observed in either animal or (rarely) human studies where subject size, health, metabolic variations, and a host of other variables can complicate the outcomes. It is generally observed in in vitro assay studies that essentially all nanoparticulates exhibit recognizable proinflammatory response. While this has little bearing on more chronic responses or the relationship to carcinogenic or mutagenic occurrences, it provides some sense of caution when processing nanoparticulates or handling them in the manufacturing arena. Figure 1 illustrates some typical examples of cytotoxicity assays for a wide range of nanoparticulate and nanoparticulate aggregate materials. In Fig. 1a, it is observed that TiO_2 allotropes are not particularly cytotoxic to human macrophage cells, while other nanoparticulates exhibit considerable cytotoxicity relative to controls. In contrast, TiO_2 allotropes and other nanoparticulate materials all exhibit cytotoxic response for human (lung) epithelial cell line exposure over a 48 h period as shown in Fig. 1b. Figure 1c illustrates a comparison of cytotoxic responses expressed as human epithelial cell death and reactive oxygen species (ROS) generation for TiO_2, soots, asbestos, and carbon nanotube materials.

References

Aguilar Z (2012) Nanomaterials for medical applications. Elsevier, New York

Alleman JE, Mossman BT (1997) Asbestos revisited. Sci Am 7:70–78

DeVolder MFL, Towfick JH, Boughman RH, Hart AJ (2013) Carbon nanotubes: present and future commercial applications. Science 339:535–539 (with supplementary online material)

Garza KM, Soto KF, Murr LE (2008) Cytotoxicity and reactive oxygen species generation from aggregated carbon and carbonaceous nanoparticulate materials. Int J Nanomedicine 3(1):1–12

Grady BP (2011) Carbon nanotube-polymer composites: manufacture, properties and applications. Wiley, New York

Koch CC (1997) Synthesis of nanostructured materials by ball milling: problems and opportunities. Nanostruct Mater 9(1–8):13–22

Koch CC, Youssef KM, Scattergood RO, Murty KL (2005) Breakthroughs in optimization of mechanical properties of nanostructured metals and alloys. Adv Eng Mater 7(9):787–794

Koratkor N (2013) Graphene in composite materials. DEstech Publishers, Lancaster

Murr LE (2012) Soot: structure, composition and health effects, Chap. 1. In: Paul MC (ed) Soot: sources, formation and health effects. Nova Science Publishers, New York, pp 1–63

Murr LE, Soto KF, Garza KM (2008) Health hazards of manufactured natural environmental and other anthropogenic atmospheric nanoparticulate materials, Chap. 1. In: Ahmed W, Ali N,

Ochsner A (eds) Biomaterials and biomedical engineering. Trans Tech Publications, Zurich, pp 1–56
Nel A, Xia T, Madder L et al (2006) Toxic potential of materials at the nanolevel. Science 311(5761):622–627
Pope CA III, Dockery DW (2006) Health effects of fine particulate air pollution: lines that connect. J Air Waste Manag Assoc 56:709–742
Reza MJ, Jamaati R, Dutkiewicz J, Szpunar JA (2013) Investigation of nanostructured aluminum/copper composite produced by accumulative roll bonding and folding process. Mater Design 51:274–279
Service R (1998) Nanotubes: the next asbestos. Science 281:941–942
Soto K, Garza KM, Murr LE (2007) Toxic effects of aggregated nanomaterials. Acta Biomater 3:351–358
Taylor R, Coulombe S, Otanicar T, Phelan P, Gunawan A, Lu W, Rosengarten G, Prasher R, Tyagi H (2013) Small particles, big impacts: a review of the diverse applications of nanofluids. Appl Phys Rev 113:011301-1–011301-19
Tran C-D, Lucas S, Phillips DG, Randeniya LK, Baughman RH, Tran-Cong T (2011) Manufacturing polymer/carbon nanotube composite using a novel direct process. Nanotechnology 22:1–9
Ward E (2005) Stomach cancer. In: Levy BS (ed) Preventing occupational disease and injury, 2nd edn. American Public Health Association, Washington, DC, p 461
Weertman JR (2012) Retaining the nano in nanocrystalline alloys. Science 337:921–922
Whang SH (ed) (2011) Nanostructured metals and alloys: processing, microstructure, mechanical properties and applications. Woodhead Publishing, Cambridge, UK
Zehetbauer MJ, Zhu YT (eds) (2009) Bulk nanostructured materials. Wiley-VCH, Weinheim

Part XIII

Extreme Deformation of Engineering Materials

Ballistic and Hypervelocity Impact and Penetration

Contents

Abstract

Fundamentals of plane shock wave loading are described in relation to spherical shock loading which is the characteristic of ballistic and hypervelocity impact and penetration of materials, particularly metals and alloys. The evolution of microstructure with shock pressure and especially shock-wave-induced twinning and the role of stacking-fault free energy in fcc metals and alloys are discussed along with critical shock pressure for twinning and the twinning-microband microstructure transition for oblique shock loading and spherical shock associated with impact crater production and subsequent penetration of targets. This includes the role played by dynamic recrystallization (DRX) in solid-state flow which facilitates crater formation. The penetration of rods in contrast to spherical projectiles in thick targets and plug formation and ejection by adiabatic shear band formation is described along with the fundamental, dynamic recrystallization (DRX) microstructure which characterizes such shear bands. Shaped charge and explosively formed penetrator structures and performance as penetrators in thick target materials are described. The role of DRX in these processes is also discussed. Finally, rail erosion characteristic of DRX in railgun operation is illustrated briefly as a related example of extreme plastic deformation along with the concept of explosively driven magnetic flux generation.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_49

Introduction: Plane Shock Fundamentals

Any impact will produce a shock wave or pulse such that the pulse velocity will increase with an increase in the shock pressure:

$$\frac{\partial^2 P}{\partial^2 U_s} > o, \tag{1}$$

where P is the shock pressure and U_s is the velocity of the shock front or shock pulse propagating in the impacted material or target. For a plane shock wave, which can be created ideally by a plane plate projectile striking parallel to the surface of a plane target, both the projectile and target plate are compressed in opposite directions at the contacting surfaces. Such an event can be created by generating an explosive – plane wave generator which produces a plane deformation front which strikes an explosive main charge, subsequently generating a high-pressure plane shock which drives a flyer plate or impact (projectile) plate against a target. These conditions are implicit in the sketches provided in Fig. 1. In a crystalline target, the shock front produces dislocations or related defects in its wake, and the pressure of the shock front attenuates as it propagates through the target (Fig. 1a). This occurs as a consequence of the pressure, density, energy, and volume (P, ρ, E, V) difference before and behind the propagating plane shock front as shown in Fig. 1b.

Assuming the impacting (projectile) plate as shown schematically in Fig. 1c has an initial velocity u_o, after impact the particle velocity in the compressed region of the projectile plate is $u_o - U_{p(p)}$, where $U_{p(p)}$ is the upward particle motion. The corresponding target particle velocity ($U_{p(t)}$) is therefore

$$U_o = U_{p(p)} + U_{p(t)}, \tag{2}$$

and if the projectile and target plates are the same material, then, $U_{p(p)} = U_{p(t)}$ U_p, and

$$u_o = 2\ Up\ U_d (\text{the driver or projectile plate velocity}). \tag{3}$$

In addition, the *equation of state* (*EOS*) is a linear relationship for shock velocity, U_s, and particle velocity, U_p, for most metals and alloys:

$$U_s = C_o + SU_p \tag{4}$$

where C_o is the longitudinal or bulk sound velocity at zero pressure in the material and S is a material constant related to the so-called Grüneisen parameter (Meyers 1994). Considering these parameters, the shock pressure generated in the target for plane shock impact as shown schematically in Fig. 1c is given by

$$P_t = P_s \text{ (the shock pressure)} = \rho_t \left(C_t + S\ U_{pm}\right) U_{pm} \tag{5}$$

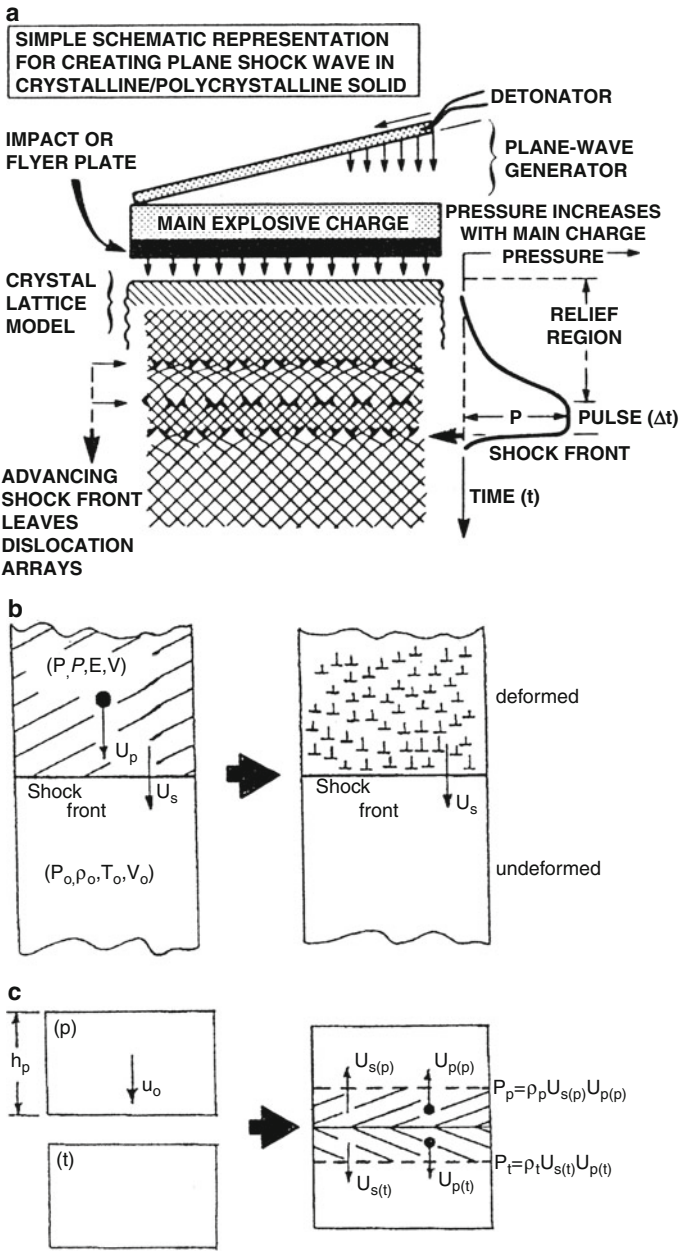


Fig. 1 Plane shock wave generation and principles schematics. (**a**) Plane wave (or line-wave) generator and plane shock production by detonation of main explosive charge (see Murr 1981). Shock-front details are illustrated. (**b**) Shock parameters and shock-front propagation. Dislocations and related microstructures form behind the shock front. (**c**) Impact of plane plate with plane target showing shock and particle velocities (After Esquivel and Murr 2006)

where the subscript t refers to the target material and U_{pm} is the modified projectile velocity as represented in Eq. 2 in the compressed region after impact (Fig. 1c) and given by Meyers (1994):

$$U_{pm} = \frac{\left[-\left(\rho_t C_t + \rho_p C_p + 2\rho_p S_p u_o\right) \pm (\Delta)^{1/2}\right]}{2\left(\rho_t S_t - \rho_p S_p\right)}, \tag{6}$$

where

$$\Delta = \left(\rho_t C_t + \rho_p C_p + 2\rho_p S_p u_o\right)^2 + 4\left(-\rho_p\right)\left(\rho_t S_t - \rho_p S_p\right)\left(C_p u_o + S_p u_o{}^2\right). \tag{7}$$

Note that in Eqs. 6 and 7, C_t and C_p refer to the corresponding bulk sound velocities, while S_t and S_p denote the corresponding Grüneisen parameters for the target and projectile materials, respectively.

In density-compatible systems (where $\rho_p = \rho_t$), the projectile or driven plate velocity, U_d, as expressed in Eq. 3 can be calibrated for any explosive composition as shown ideally in Fig. 2 for a C-4 explosive main change (Fig. 1a), considering the ratio of explosive main charge height, h_e; explosive density, ρ_e; and the corresponding height (or thickness) of the projectile (impact) or driven plate, h_d, and its corresponding density, ρ_d: $h_e\rho_e/h_d\rho_d$. Table 1 compares the properties (including the density) for a number of explosive compositions.

In a plane wave generator arrangement as shown in Fig. 1a, the duration of the shock pulse generated in the target is given by

$$\Delta t = \frac{2\,h_d}{U_s}, \tag{8}$$

where U_s is related to the shock pressure, ρ_s, for any target material and $h_d = h_p$ in Fig. 1c. As it turns out, the shock pressure, P_s, is also related to the ratio $h_e\rho_e/h_d\rho_d$, so it is possible to select a shock pressure and pulse duration (Δt) by varying the projective or driver plate thickness, h_d, along with the height of the main explosive charge, h_e, in Fig. 1a. Of course the projectile or driver plate velocity as represented in Eq. 5 (the modified projectile velocity) can be directly measured. In addition, the plane-wave-generating projectile can be launched from other energetic sources such as a gun or magnetic flyer system.

The relationship between pressure (P) and volume (V), or P-V data, is plotted as a *Rankine* or *Hugoniot* curve for many materials as illustrated for a few representative metals and alloys in Fig. 3. These plots represent the loci of P-V states which are accessible by a shock transition. That is, considering conservation of mass at the shock front requires that

$$\left(U_s - U_p\right)\rho = U_s\rho_o, \tag{9}$$

similar arguments for momentum and energy conservation are expressed as

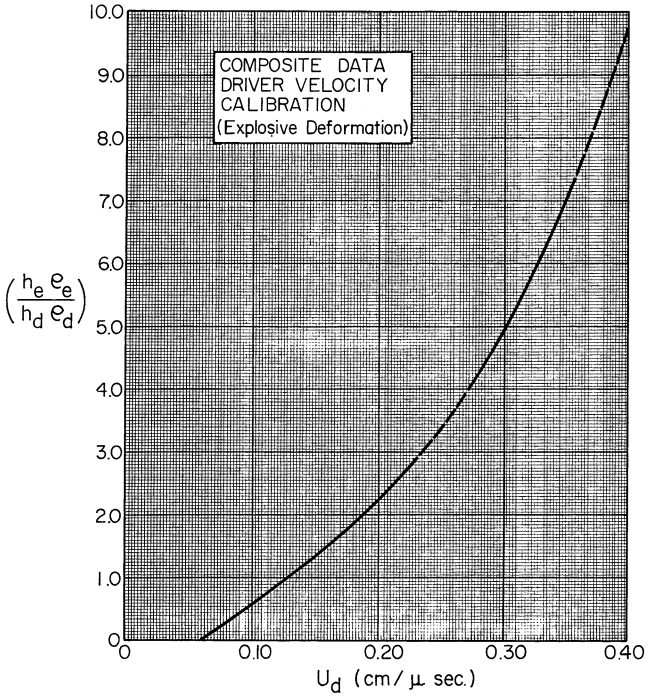


Fig. 2 Penetrator/driven plate velocity (U_d) calibration curve for composition C-4 (Table 1) main charge (Fig. 1a) in plane wave explosive deformation. Applicable for density-compatible tooling arrays as in Fig. 1a with $\rho_t = \rho_p$

$$(P - P_o) = \rho_o U_s U_p \tag{10}$$

and

$$(E - E_o) = \frac{(P + P_o)(V_o - V)}{2}. \tag{11}$$

Assume hydrostatically compressed material $V_o = 1/\rho_o$; $V_o = 1/\rho$ in the uncompressed and compressed regions in Fig. 1b, respectively. Rearranging Eqs. 9 and 11 results in

$$U_s = V_o[(P - P_o)(V_o - V)]^{1/2} \tag{12}$$

and

$$U_p = [(P - P_o)/(V_o - V)]^{1/2}. \tag{13}$$

Equations 9 through 13 are often referred to as the Rankine–Hugoniot equations. Meyers and Murr (1981) developed tables of Hugoniot P-V data through computer solutions of these equations.

Table 1 Selected explosives and properties (Murr and Hare 1988)

Explosive	Composition	Detonation velocity V_D		Density (g/cm^3)	Heat of explosion (cal/g)	$\sqrt{2E}$[d] (mm/μs)
		m/s	ft/s			
HMT (BETA)	–	9,120	29,914	1.84	1,360	3.1
PETN	Pentaerythritol tetranitrate $C_5H_8N_{12}O_4$	8,300	27,224	1.70	1,390	2.8
RDX[a]	$C_3H_6N_6O_6$	8,180	26,830	1.65	1,280	2.8
C-4	RDX/91	8,040	26,371	1.59	1,230	2.7
COMP. B	RDX/60; TNT/40	7,840	25,715	1.68	1,240	2.6
Nitroglycerine	$(C_6N_7N_3O_{11})_n$	7,700	25,256	1.60	1,600	2.6
Pentolite	PETN/50; TNT/50	7,470	24,502	1.66	1,220	2.5
Detasheet C	PETN/68	7,200	23,616	1.45	990	2.4
TNT	Trinitrotoluene – $C_7H_5N_3O_6$	6,700	21,976	1.56	1,080	2.3
DBX	RDX/21; AN/21; TNT/40 + aluminum powder	6,600	21,648	1/65	1,700	2.2
Dynamite (Hercules)	RDX/75; TNT/15	6,000	19,680	1.10	940	2.0
Dynamite[b] (low velocity)	TNT/68	4,400	14,432	0.90	625	1.5
Ammonium[c] nitrate (ANFO)	+Perlite/fuel oil (NH_4NO_3)	3,000 to 1,500	9,800	0.8–1.0	350	<1
Ammonium nitrate/8 % fuel oil (ANFO)	–	2,400	7,872	0.85	–	<1
Ammonium nitrate/15 % fuel oil (ANFO)	–	1,500	4,920	0.90	–	<1

[a]Cyclotrimethylenetrinitramine
[b]Originally dynamite was simply nitroglycerine mixed with a chemically inert material such as wood pulp to make the nitroglycerine safer to handle. As indicated here, dynamite is now applied to explosives without nitroglycerine
[c]AN = ammonium nitrate. Ammonium nitrate mixtures can be made faster (increasing V_D) by adding up to 15 % aluminum powder
[d]Gurney energy = $V_D/2.97$

The strain associated with the instability at the shock front (Fig. 1) is expressed as dynamic (true) strain:

$$\varepsilon = 4 \ln \frac{(V/V_o)}{3}, \tag{14}$$

where V/V_o is determined from Rankine–Hugoniot plots such as those in Fig. 3 at the pressure characteristic of the shock pressure. It is apparent that these strain

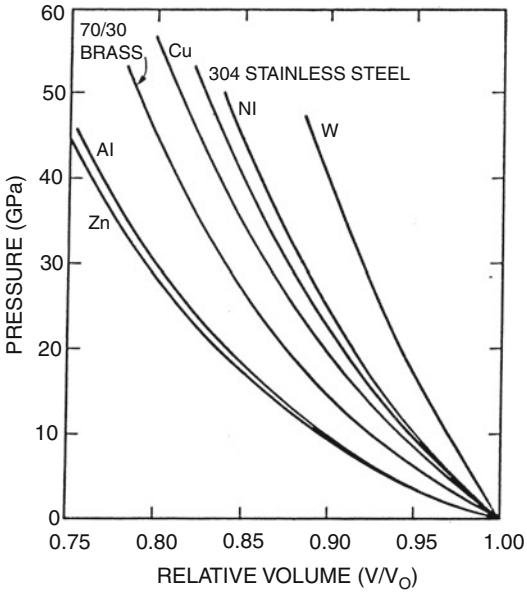


Fig. 3 Rankine–Hugoniot compressibility curves for several metals and alloys (plane shock wave propagation)

values are small in plane shock, since for Cu or Ni in Fig. 3, at a shock pressure of ~50 GPa, the corresponding strains are −0.2 for Eq. 14. The strain rates are related to the shock pulse duration and are usually $>10^6$ s^{-1}. These conditions impose deformation and deformation-related microstructures behind the shock front as illustrated typically in Fig. 4 for plane shock-loaded nickel at increasing peak pressures, with a constant shock pulse duration (Δt) of 2 μs. Figure 4 shows a transformation or evolution of dislocation density and associated dislocation cell size with pressure up to a critical pressure at which deformation twinning occurs. This critical twinning pressure is ~30 GPa for Ni in Fig. 4, but decreases with decreasing stacking-fault free energy as illustrated in Fig. 5 for other metals and alloys.

The evolution of microstructure in plane shock deformation implicit in Fig. 4 can be expressed in an extended Hall–Petch relationship in the form

$$\sigma_y = \sigma_o + K/\sqrt{D} + K'/d + K''/\sqrt{\Delta} \tag{15}$$

where K, K′, and K″ are material constants, D is the grain size, d is the dislocation cell size or spacing, and Δ is the deformation twin spacing. In Fig. 4a, b the residual yield stress becomes

$$\sigma_y = \sigma_o + K/\sqrt{D} + K'/d, \tag{16}$$

where d decreases from ~0.4 μm in Fig. 4a to ~0.2 μm in Fig. 4b. Correspondingly, in Fig. 4c, d the residual yield stress becomes

$$\sigma_y = \sigma_o + K/\sqrt{D} + K''/\sqrt{\Delta} + K'''' \sqrt{\rho} \tag{17}$$

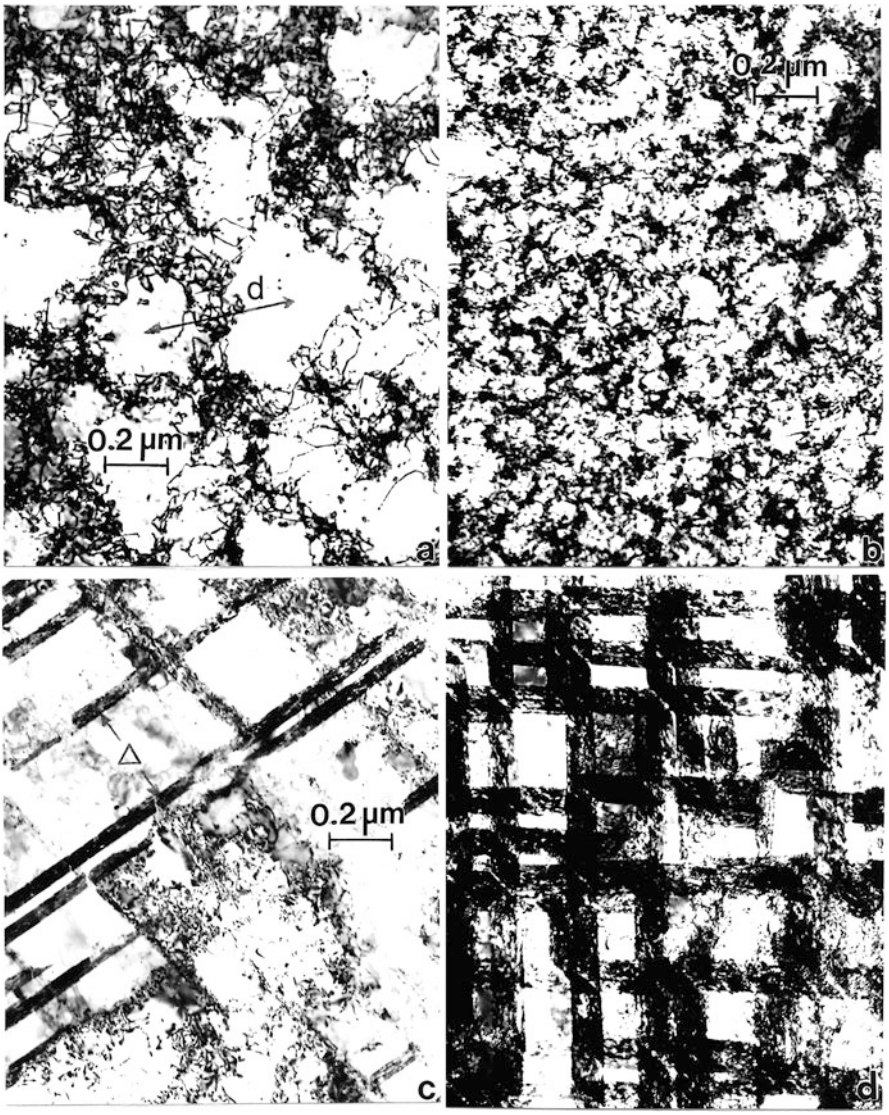


Fig. 4 Plane shock wave microstructures in nickel: TEM images at increasing shock pressures noted. (**a**) 15 GPa; (**b**) 25 GPa; (**c**) 30 GPa; (**d**) 45 GPa. Note magnification in (**d**) is same as (**a**) to (**c**) (After Murr 2012)

where K''' is another material constant and ρ is the dislocation density. The inter-twin spacing, Δ, in Eq. 17 decreases from ~0.3 μm in Fig. 4c to ~0.15 μm in Fig. 4d. The corresponding microindentation hardness corresponding to Fig. 4a, b can be written

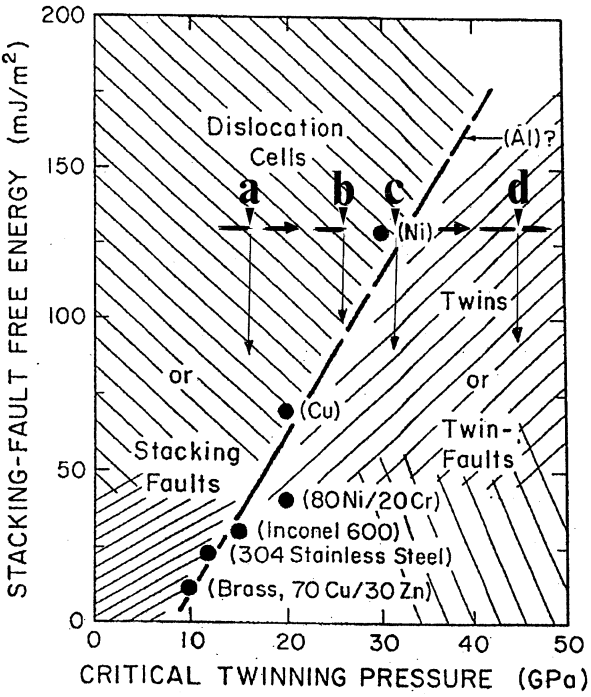


Fig. 5 Stacking-fault free energy (room temperature) versus critical twinning pressure for plane shock for fcc metals and alloys. The Ni line progression (**a**–**d**) shown corresponds to Fig. 4a–d microstructures (From Murr 2012)

$$H = H_o + K/\sqrt{d} + K'/d, \tag{18}$$

and corresponding to Fig. 4c, d

$$H = H_o + K/\sqrt{D} + K''/\sqrt{\Delta} + K'''' \sqrt{\rho}. \tag{19}$$

From Eqs. 18 and 19, it can be observed that the residual hardness for plane shock-loaded metals and alloys would be expected to continue to increase with shock pressure. This is generally observed in Fig. 6 where residual microindentation hardness is plotted against the square root of the shock pressure, P, for a number of fcc and bcc metals and alloys. Figure 7 compares the residual microindentation hardness measurements for plane shock-loaded and cold-rolled Ni and 304 stainless steel, representing a stacking-fault free energy of ~128 mJ/m^2 versus 22 mJ/m^2, respectively, against the measured true strain.

Figure 8 compares the stress–strain diagrams for plane shock-loaded 70/30 (Cu/Zn) brass and cold-rolled brass (Fig. 8a, b) along with a comparison of plots of ultimate tensile strength (UTS) and residual microindentation hardness measurements for 70/30 brass and 304 stainless steel deformed by plane shock loading and cold rolling. Figures 9 and 10 compare plane shock-loaded 70/30 brass and 304 stainless steel microstructures. The 70/30 brass (Fig. 9) with a stacking-fault free energy of ~15 mJ/m^2 (Murr 1975) exhibits a planar dislocation structure with

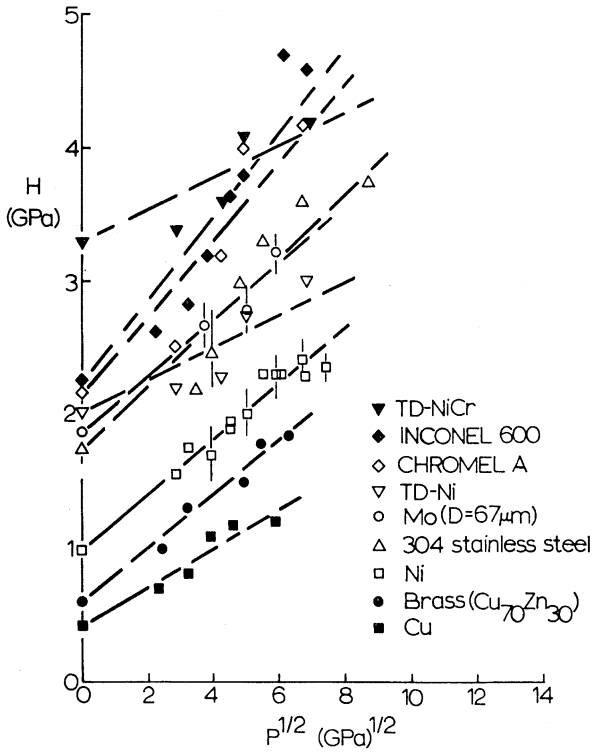


Fig. 6 Vickers microindentation hardness (H) versus square root of peak shock pressure (P) for plane shock-loaded metals and alloys

no recognizable stacking faults at 4 GPa shock pressure, while the 304 stainless steel, with a stacking-fault free energy of ~22 mJ/m^2, exhibits extensive stacking-fault microstructures at 15 GPa peak shock pressure in Fig. 10.

Figure 11 shows tantalum (Fig. 11a) after dynamic (10^3 s^{-1} strain rate) compression (Fig. 11b) compared to the same tantalum dynamically compressed and plane shock-loaded at 45 GPa peak pressure (Fig. 11c). Figure 11d shows a TEM image of the deformation twins shown in the compressed grains of Fig. 11c.

In contrast to compressive loading, shock loading, specifically plane shock loading, does not cause any grain distortion (Fig. 11c), at least at shock pressures below about 80 GPa. However, at very high peak pressures, there is some distortion combined with dynamic recrystallization. In addition, at plane shock pressures above 100 GPa, there is a trailing thermal pulse which can alter the microstructure and cause residual softening. This is illustrated in Fig. 12. Figure 12a shows residual microindentation hardness measurements for Ni and 304 stainless steel to exhibit high-pressure softening, while Fig. 12b illustrates an unusual cell-like microstructure for their high-pressure shock loading of 304 stainless steel at 120 GPa. This microstructure is notably distinct from the low-pressure stacking-fault microstructure illustrated in Fig. 10. Indeed, it can be noted in considering Figs. 4, 5, 6, 7, 8, 9, 10, 11, and 12 in retrospect that plane shock deformation does not induce severe plastic deformation characterized by dynamic recrystallization

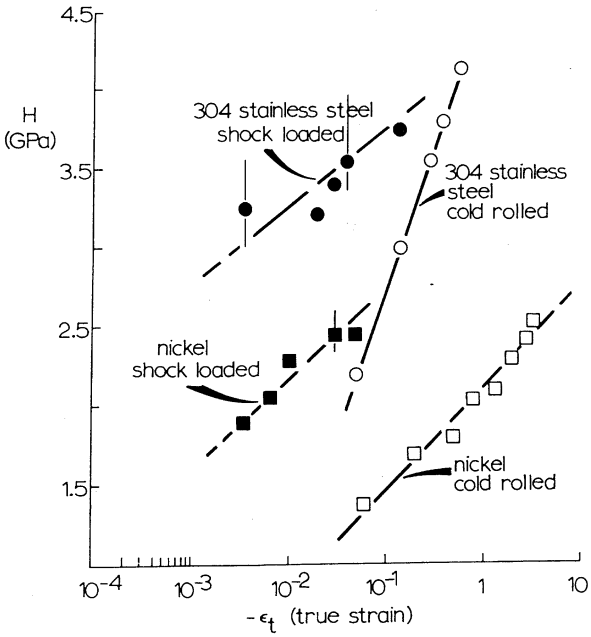


Fig. 7 Comparison of residual microindentation (Vickers) hardness versus true strain for plane shock-loaded and cold-rolled nickel and type 304 stainless steel

except at extreme shock pressures as implicit in Fig. 12. Consequently, shock hardening has been a useful tool in surface treatment of very large metal or alloy structures because there is little or no distortion associated with the more than a doubling of the base hardness (Fig. 6) near the shocked surface (Murr 1988).

Spherical Shock: Ballistic and Hypervelocity Impact and Crater Formation

It is apparent from Fig. 1 that geometry dictates the conditions of impact: projectile impacting a target. For a spherical or ideally spherical projectile, the geometry at impact ($t = 0$) will differ from plane impact, and for $t > 0$, the shock wave produced in the impacted target will follow the geometry of the projectile. Furthermore, the impacting projectile as illustrated in Fig. 13 will excavate a crater in the target by creating a fluidlike flow regime at the projectile/target interface. This occurs by creating a severe plastic deformation zone at the surface of the target and projectile so that target material flows from the target as the projectile penetrates. In this hydrodynamic regime, the steady-state pressure along the line of impact is expressed by the *Bernoulli equation*:

$$P_B = \frac{\left[\rho_t \rho_p / \left(\rho_t^{1/2} + \rho_p^{1/2}\right)^2\right] u_o^2}{2}, \tag{20}$$

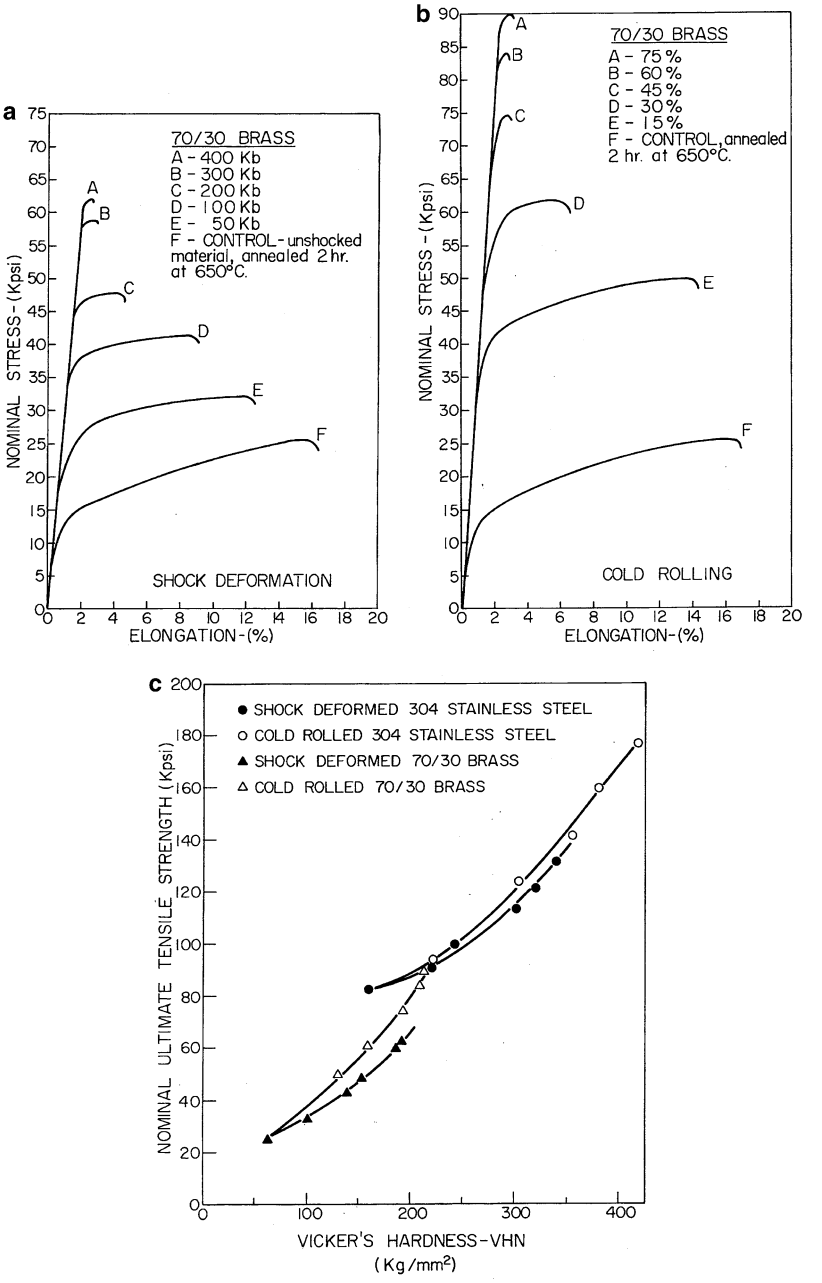


Fig. 8 Residual mechanical properties for plane shock-loaded and cold-rolled brass and stainless steel. (**a**) Stress–strain diagrams for shock-loaded 70/30 brass. (**b**) Stress–strain diagrams for cold-rolled 70/30 brass. (**c**) UTS versus residual hardness for shock-loaded and cold-rolled 70/30 brass and type 304 stainless steel. Note 100 kb (kilobar) = 10 GPa; 100 VHN (kg/mm^2) = 1 GPa

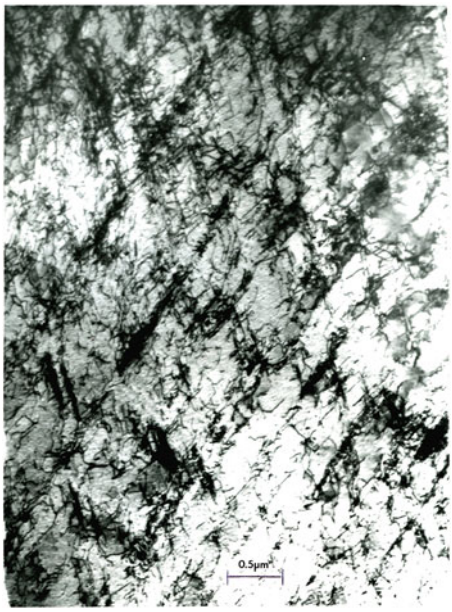


Fig. 9 TEM image of planar dislocation arrays in plane shock-loaded 70/30 brass at 5 GPa peak pressure. The surface orientation is (112)

where ρ_t and ρ_p are the target and projectile densities and u_o is the projectile velocity at impact ($t = 0$). When the projectile and target are the same ($\rho_t = \rho_p$), Eq. 20 becomes

$$P_B = \frac{\rho u_o^2}{8}. \tag{21}$$

Equations 20 and 21 are characteristic of the axial pressure at the crater bottom at steady state. The shock pressure is attenuated from $t = 0$ to $t > 0$ as given at the point of impact by Eq. 6. It can be observed on comparing Eqs. 6 and 20 that the instantaneous shock pressure is notably greater than P_B.

The depth of penetration for any penetrator having a length ℓ_o is given by

$$p \propto \ell_o \left(\rho_p/\rho_t\right)^{1/2} \tag{22}$$

where for an ideally spherical penetrator (Fig. 13a) $\ell_o \cong d_p$. Correspondingly, for $\rho_p = \rho_t$, Eq. 22 becomes

$$p \propto d_p. \tag{23}$$

While Eq. 22 implies that the penetrator velocity at impact, u_o, is not a consideration for penetration, this is not entirely true, although at hypervelocity, generally described as $u_o \geqslant 5$ km/s, the crater geometry can be expressed by Watts et al. (1993):

Fig. 10 TEM image of stacking faults in plane shock-loaded type 304 stainless steel at 15 GPa

$$D_c/d_p = \left[1.1\left(\rho_p/\rho_t\right)^{1/3}(\rho_t/\sigma_t)^{1/3}\left(C_t/C_p\right)^{1/3}{u_o}^{2/3}\right] \Big/ \left[1+\left(\rho_p/\rho_t\right)^{1/2}\right]^{2/3}, \tag{24}$$

where D_c and d_p are the crater diameter and penetrator diameter as shown in Fig. 13a, σ_t is the target yield strength, and C_t and C_p are the target and penetrator sound velocities. Again, in the ideal case where $\rho_t = \rho_p$ and $C_t = C_p$, Eq. 24 is simplified to

$$D_c/d_p = \zeta\left(u_o/\sqrt{\sigma_t}\right)^{2/3}, \tag{25}$$

where ζ is a constant. In addition, if σ_t is expressed by a Hall–Petch-type relationship as in Eqs. 15, 16, and 17, it is apparent that the microstructure for the target will influence the crater parameters. Since the hardness is generally related to the yield stress for metals and alloys,

$$H_t \cong 3\sigma_t, \tag{26}$$

and H_t can be substituted into Eq. 25 for impact and crater formation at hypervelocity. Watts et al. (1993) have also shown that

$$p/D_c = \zeta^*(u_o/\sigma_t)^{1/3} \tag{27}$$

for $u_o < 5$ km/s (u_o below hypervelocity) and

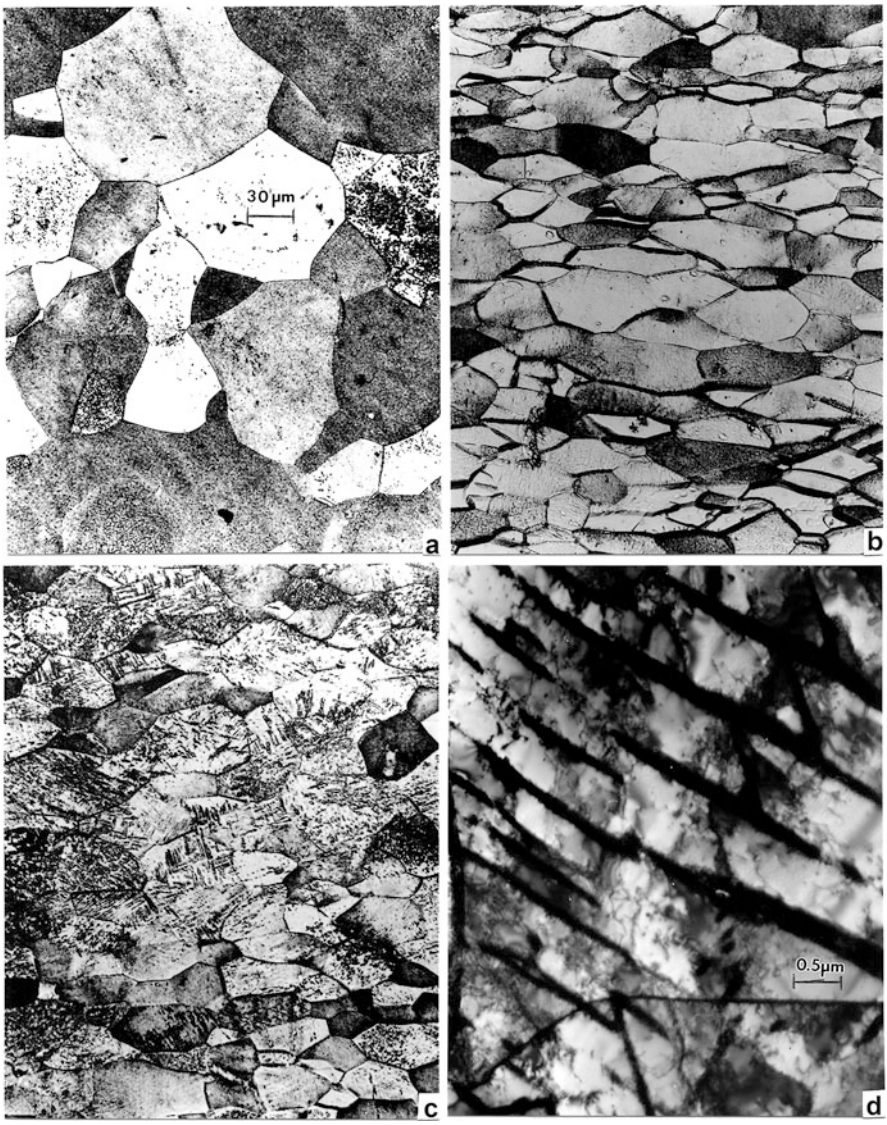


Fig. 11 Deformation of tantalum. (**a**) Original polycrystalline stock. (**b**) After dynamic compression. Note grain distortion in comparison to (**a**). (**c**) Dynamic compression followed by plane shock at 45 GPa. (**d**) TEM image of deformation twins in (**c**)

$$p/D_c = \zeta^* \left(u_o/\sqrt{\sigma_t}\right)^{2/3} \tag{28}$$

for $u_o > 5$ km/s (u_o above hypervelocity), where ζ^* is some constant as implicit in Eq. 25. It has been shown experimentally for a variety of impact parameters, i.e., different penetrator densities and target densities, that $p/D_c \sim 0.5$ for $u_o > 5$ km/s

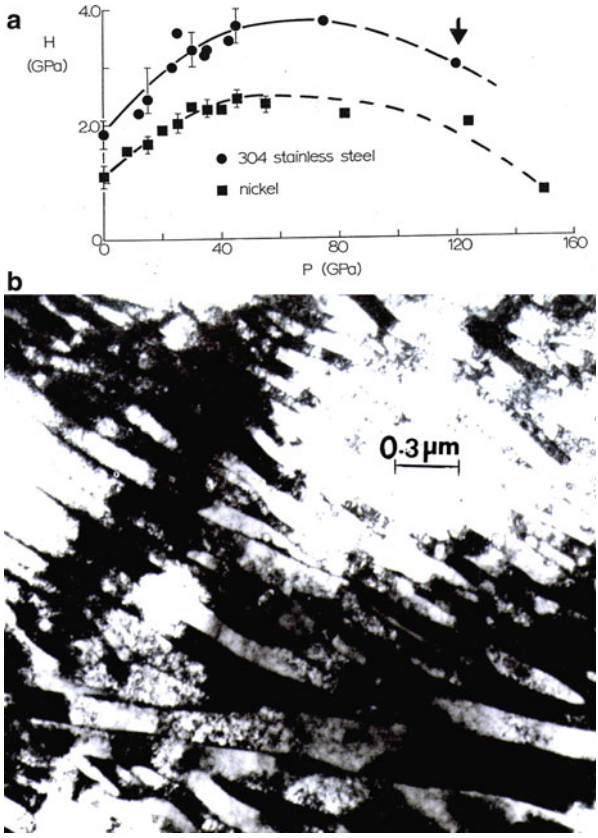


Fig. 12 Very high-pressure plane shock loading effects in Ni and 304 stainless steel. (**a**) Hardness versus shock pressure (corresponding to 2 μs shock pulse). (**b**) TEM image of shock-induced high-pressure cell microstructure in 304 stainless steel (120 GPa)

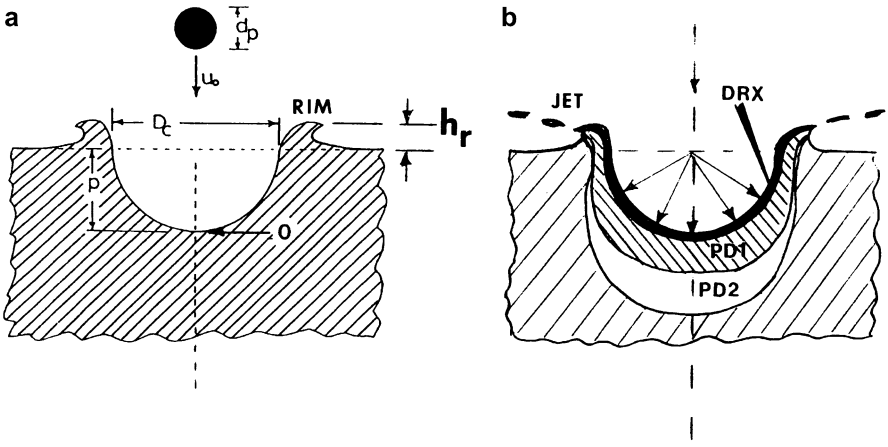


Fig. 13 Schematic views for impact crater formation (**a**) and creation of deformation zones by spherical shock (**b**). Plastic deformation zones below the dynamic recrystallization zone (DRX) are denoted PD1 and PD2 and discussed in the text

for projectiles of lighter density, such as glass or characteristic chondritic (meteoric) space particles (where $\rho \leqslant 2.2$ g/cm^3). However, between $\rho \leqslant 2.3$ and > 15 g/cm^3, p/D_c has been observed to increase from 0.5 to 1.2 (Valerio-Flores et al. 2004). Correspondingly, $D_c/d_p \sim 3$ for $u_o > 5$ km/s (Quinones et al. 1995). These conditions generally apply for infinite or semi-infinite targets or for projectiles impacting normal to the target surface as implicit in Fig. 13a. However, for finite target thicknesses, and for very thin targets, spallation or penetration can occur. For very thin target plates, penetration will occur before spallation at some critical velocity relative to the thickness, a condition referred to as the ballistic limit, a projectile velocity above which penetration will occur for a target thickness, t_t. However, this is a very complex issue and will depend upon not only the target thickness, t_t, but also the projectile diameter, d_p, the angle of incidence (or deviation from the normal as shown in Fig. 13a). Formulations for thin targets often appear in empirical forms based on experimental testing, e.g. (Christiansen 2009),

$$d_p \cong \left(\frac{t_t \, H_t * 1/4 \, \left(\rho_t/\rho_p\right)^{1/2}}{5.24 \, k \, \left(u_o \cos\theta/C_t\right)^{2/3}} \right)^{0.95}, \tag{29}$$

which covers the hypervelocity range ($u_o > 5$ km/s); H_t* is the Brinell hardness of the target plate, and K is a damage parameter which is equal to ~1.8 for penetration, between 1.8 and 3 for spallation or spall detachment, and >3 for cratering and no spallation or penetration. d_p in Eq. 29 is the projectile diameter as shown in Fig. 13a. C_t in Eq. 29 is the target sound velocity. An empirical equation for the velocity of a projectile characteristic of the *ballistic limit* velocity (u_b) has been developed by Reid and Wen (2000) in the form

$$u_b = \frac{\pi\gamma\sqrt{\rho_t \, \sigma_a}}{4m} \quad d_p{}^2 \, t_t \left(1 + \sqrt{1 + \frac{8m}{\pi\gamma^2 \, \rho_t \, d_p^2 \, t_t}} \right), \tag{30}$$

where γ is an experimental constant, σ_c is the static linear elastic compression limit (a material parameter related to hardness or yield strength), and m is the projectile mass. Equation 30 assumes normal impact of the projectile; that is, Θ in Eq. 29 is zero. The penetration depth, p, is also often represented by the Charters–Summers expression (Charters 1960):

$$p = \left(\frac{81}{4\pi} \left(\frac{\rho_p}{\rho_t} \right) \frac{E}{S_t} \right)^{1/3} \tag{31}$$

where E is the kinetic energy of the projectile:

$$E = m_p u_o{}^2/2, \tag{32}$$

m_p in Eq. 32 is the projectile mass, and S_t in Eq. 31 is the target strength. Using Eq. 32 it is possible to estimate the impact velocity (u'_o) at which the target (or crater volume) will melt:

$$E' = m_p u'_o{}^2/2 = \int_{T_o}^{T_m} M_t\left(c_p dT + L\right), \tag{33}$$

where M_t is the target mass, C_p is the specific heat capacity at constant pressure, L is the specific latent heat, and T_o and T_m are the room temperature and melt temperature for the target material. Correspondingly, the projectile velocity for melting becomes

$$u'_o = \sqrt{\frac{2}{m_p}\left(\int_{T_o}^{T_m} M_t\left(c_p\, dT + L\right)\right)} \tag{34}$$

The projectile can similarly melt and at hypervelocity the projectile can, in some cases, vaporize. In addition, projectiles are also observed to fragment on impact, and this fragmentation increases for increasing impact velocity for a variety of projectile/target systems (and densities: ρ_p/ρ_t). In fact, there exists a threshold velocity at which projectile fragmentation occurs: a fragmentation onset velocity, reported to be >2 km/s (Grady and Kipp 1998). Recent work by Hernandez et al. (2006) showed the fragmentation onset velocity for spherical stainless steel projectiles impacting a variety of metal targets between 0.5 and ~5 km/s to be ~0.7 km/s, with fragment size decreasing with increasing impact velocity. When extrapolated to ~6 km/s, the fragment size tended to be zero, indicative of projectile vaporization. Figure 14 illustrates a computer simulation for a 3.18 mm diameter stainless steel sphere impacting a plane nickel target at 3.4 km/s. This 3D section simulation also shows the half-section view of the nickel target crater development as well.

Because of projectile fragmentation and vaporization at hypervelocity, Whipple (1947) proposed that in space, adequate projection from hypervelocity space debris and micrometeorite impact (which can range from <5 km/s to 90 km/s and averages ~20 km/s) can be provided by a series of thin, spaced layers (up to 5) because fragmented projectile debris would be trapped in what are called "bumper layers." Of course this is only true for $u_o > 5$ km/s. For slower-moving projectiles, especially those characteristic of ballistic regimes, very thick target (armor) is required, especially for rod penetrators which have considerable momentum and are not fragmented at velocities, $u_o < 3$ km/s. These will be described later in this chapter.

In retrospect, especially on examining Eqs. 24 through 34, it can be observed that hypervelocity impact cratering can be influenced by the target properties which in turn are dependent upon the initial microstructure or the evolution of microstructure as the shock wave propagates from the point of impact. In addition to projectile fragmentation onset and the conditions for achieving the ballistic limit and target penetration, strong shock wave pressures can reflect from the back side of a finite

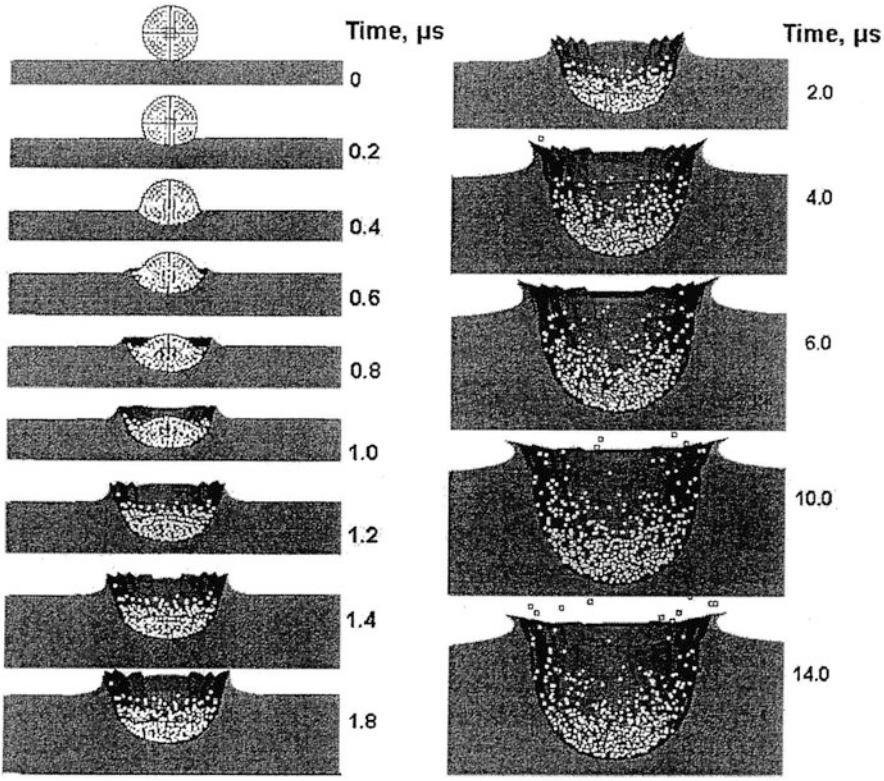


Fig. 14 3D (section) computer-simulated impact crater sequence for a 3.18 mm diameter stainless steel projectile impacting a thick Ni target at $u_o = 3.4$ km/s. Times from impact (at $t = 0$) are shown. Target thicknesses are reduced to accommodate the sequence and are not representative of the semi-infinite target (After Hernandez et al. 2006)

thickness target, and when sufficiently high, the interaction of this reflected pressure wave with the attenuating forward shock pulse can create a specific region matching the spherical wave geometry which is in tension as a consequence of the wave interaction. This equivalent stress can exceed the target material fracture stress and create a spall surface which at a sufficiently high pressure (and correspondingly high impact velocity, u_o) can become detached and ejected from the back side of the target. In some situations, spall detachment can occur prior to penetration.

Figure 15 provides an overview for crater formation, microstructure development, and evolution in the target below the crater bottom, the cratering mechanism (s), and spall formation.

Referring retrospectively to Fig. 13, it can be noted that in addition to projectile melting and simultaneous excavation of the target to form the crater, crater material can flow outward by a thin region of dynamically recrystallized (DRX) grains. This outflow of projectile and target creates a crater rim which ejects or jets

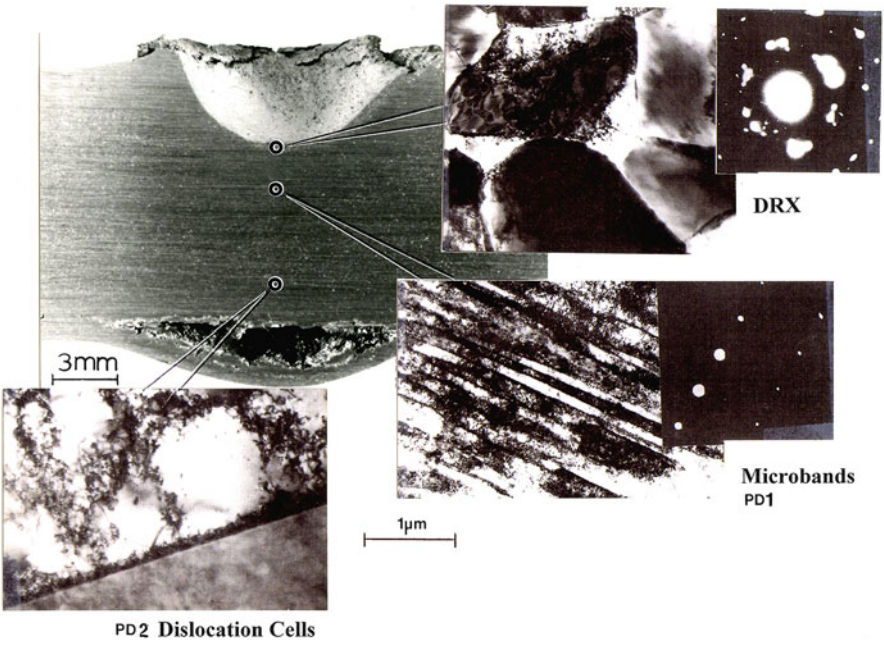


Fig. 15 Impact crater formed in Cu target by 3.2 mm diameter Al projectile at 6.01 km/s velocity showing rear-surface spall and TEM microstructure images corresponding to deformation zones shown in Fig. 13b. The selected-area electron diffraction pattern insert showing microbands represents a (112) grain surface orientation (From Murr 2012)

particles because of high resolved tensile stresses along the rim. At high stresses created by hypervelocity projectile impact as in Fig. 15, the rim can crack along with spall formation as shown. The severe plastic deformation zone or DRX zone at the crater wall is followed by a plastic deformation zone indicated by PD1 in both Figs. 13b and 15. In Fig. 15, this PD1 zone is characterized by microbands which, like deformation twins in plane shock loading, are formed at some critical shock pressure and are oriented in (or coincident with) the fcc {111} planes. The PD1 zone in Figs. 13b and 15 is followed by a reduced pressure zone which is characterized by dislocation cell structures. This microstructure evolution can be more systematically observed in the continuous optical microscope views provided in Fig. 16. Here the edge of the microband microstructure is observed to be between 1.5 and 2.1 mm from the crater wall.

Figure 17 compares deformation twin formation in plane shocked Ni at 30 GPa in contrast to microbands formed in Ni impacted by a stainless steel projectile ($\rho = 8.9$ g/cm^3) at 3.5 km/s and a corresponding shock peak shock pressure of 105 GPa. This deformation microstructure variation for plane shock versus spherical shock is illustrated more fundamentally in Fig. 18. Figure 17a shows a microtwin or twin-fault coincident with the $(\bar{1}11)$ plane which is perpendicular to the (110) grain surface reference shown. Correspondingly, Fig. 18b

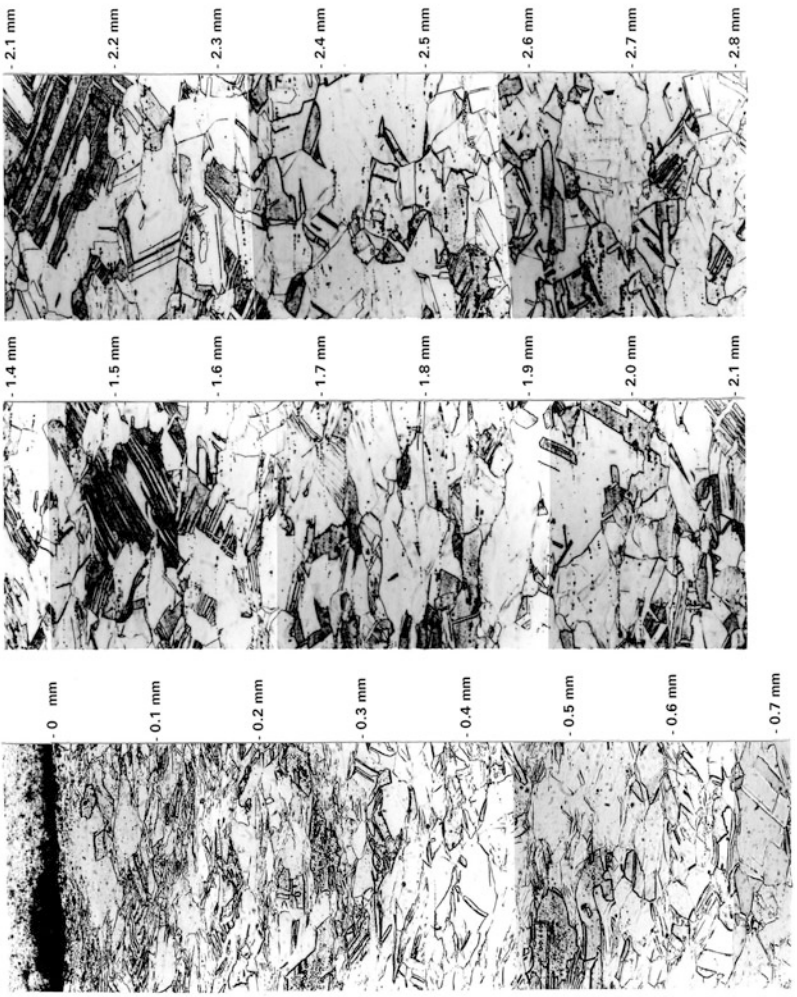


Fig. 16 Optical microscope image sequence showing microstructure sequence below the impact crater in Fig. 15 along the line of normal impact (Fig. 13a). Distance from crater wall is shown in mm

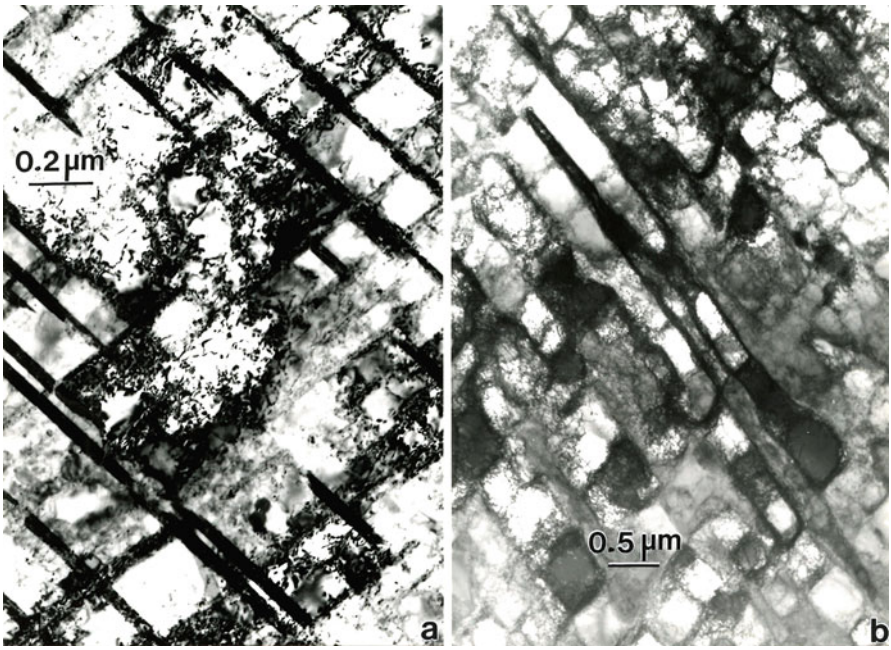


Fig. 17 Comparison of plane shock deformation twin microstructure (**a**) with spherical shock microband microstructure (**b**) in Ni. (**a**) Corresponds to a plane shock pressure of 30 GPa. (**b**) Corresponds to stainless steel 3.2 mm diameter projectile impact at 3.5 km/s forming a crater as in Fig. 13b

illustrates microband formation by dislocation double walls which are somewhat coincident with the $(\bar{1}11)$ planes. While the twin boundaries are specifically $(\bar{1}11)$ planes in Fig. 18a and have a mirror symmetry misorientation angle Θ, with the matrix, the microbands have only a small misorientation ($\sim 2^{\circ}$) relative to the fcc crystal matrix planes as illustrated in Fig. 18b.

Because the microstructure changes below the crater bottom as illustrated in Figs. 15 and 16, the corresponding residual microindentation hardness would be expected to reflect this change either along the line of impact at the crater bottom or along any radial line from the crater wall into the target material. This is illustrated for a crater in copper as shown in Fig. 19. The crater in Fig. 19a resulted from a soda-lime glass (slg) projectile 3.2 mm in diameter, impacting at a velocity $u_o = 5.8$ km/s. Figure 19b shows the microhardness profile along the impact axis for the impact projectile (slg) corresponding to Fig. 19a and a stainless steel projectile impacting the same copper target but at $u_o = 2$ km/s. The hypervelocity slg projectile ($u_o = 5.8$ km/s) created a DRX zone at the crater wall which allowed the indenter to record the microindentation hardness as a dramatic reduction shown in Fig. 19b, while both hardness profiles in Fig. 19b show a declining hardness extending from the crater bottom. This corresponds to the microstructure evolution illustrated in Figs. 15 and 16. The actual microstructure evolution corresponding to

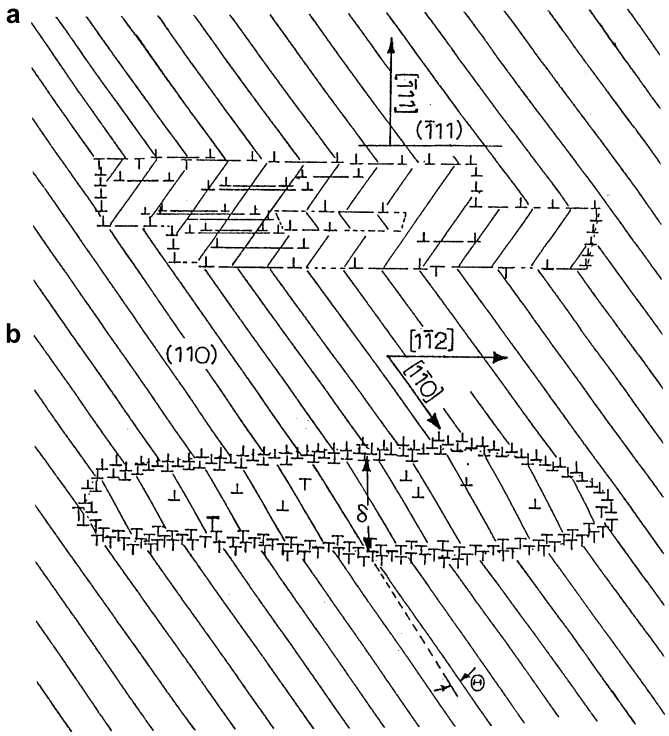


Fig. 18 Schematic views comparing deformation twin (or twin-fault) features (**a**) with microband formation (**b**) in (110) fcc grain. Both features are aligned in [1$\bar{1}$2] direction. The twin planes and associated stacking faults in (**a**) are coincident with ($\bar{1}$11) planes at 90° to the (110) grain surface. Microband width is denoted δ, Θ denotes misorientation

Fig. 19a designated in the sequence marked a to d in Fig. 19b is shown in Fig. 20. This sequence of microstructure evolution closely matches that for impacted copper in Fig. 15.

Just as deformation twin volume fraction increases with increasing grain size as illustrated previously in ▶ Fig. 9 in chapter "Strengthening by Crystal Imperfections," microband volume fraction also increases with increasing grain size as illustrated in Fig. 21. The deformation twin formation relates to plane shock, while the microband formation relates to spherical shock deformation as implicit in Fig. 13b.

The evolution of microstructure with impact crater formation at increasing projectile impact velocities can be illustrated by microindentation hardness mapping around the crater wall in recovered or residual impact crater half sections. This is illustrated in Fig. 22 for a 3.2 mm diameter aluminum 1,100 projectiles ($\rho = 2.7$ g/cm^3) impacting a copper target plate at velocities ranging from 1.08 to 6.01 km/s, illustrated previously in Fig. 15. Figure 22a–d shows the normal and half-section impact crater views, while Fig. 22e–h shows the corresponding microindentation

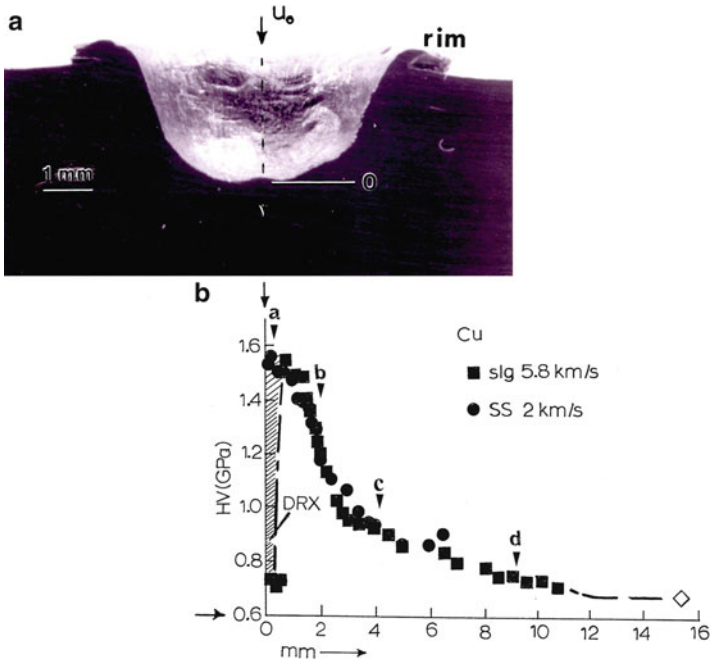


Fig. 19 (**a**) Impact crater in Cu target formed by soda-lime glass (3.2 mm diameter) projectile at $u_o = 5.8$ km/s. (**b**) shows the microindentation hardness profile measured for the crater in (**a**) along with data for another crater in a Cu target formed by a stainless steel projectile (3.2 mm diameter) at a velocity of 2 km/s. Distance from the crater bottom is measured in mm (From Murr 2012)

hardness mappings for the crater half sections, with Vickers microindentation zones illustrated in the color key shown. Only Fig. 22g, h indicates a thin DRX zone at the crater wall, while Fig. 22h can be compared to the microstructure zones implicit in Figs. 15 and 16.

Figure 23a–d shows the 2D computer simulations for residual yield stress mapped around the crater walls which correspond to Fig. 22e–h: recalling that $\sigma_y \cong H/3$. Note the incipient spall crack simulated at the arrow in Fig. 23c as well as the soft (melted) projectile remnants in Fig. 23a, b. A very thin DRX zone is simulated at the crater wall in Fig. 23c, d. This DRX zone is expanded with increasing impact pressure which is simulated in Fig. 23f, g, in contrast to the spall simulation in Fig. 23d, shown for comparison in Fig. 23e. The detachment of the spall is also shown in Fig. 23f, g, while Fig. 23g also shows penetration of the target by the projectile as well.

Simulations such as those shown in Figs. 14 and 23 utilize appropriate algorithms which connect the yield stress regime with the associated strain, strain rate, and temperature. Functionally, this involves

$$\sigma_y = f\,(\varepsilon, \dot{\varepsilon}, T). \tag{35}$$

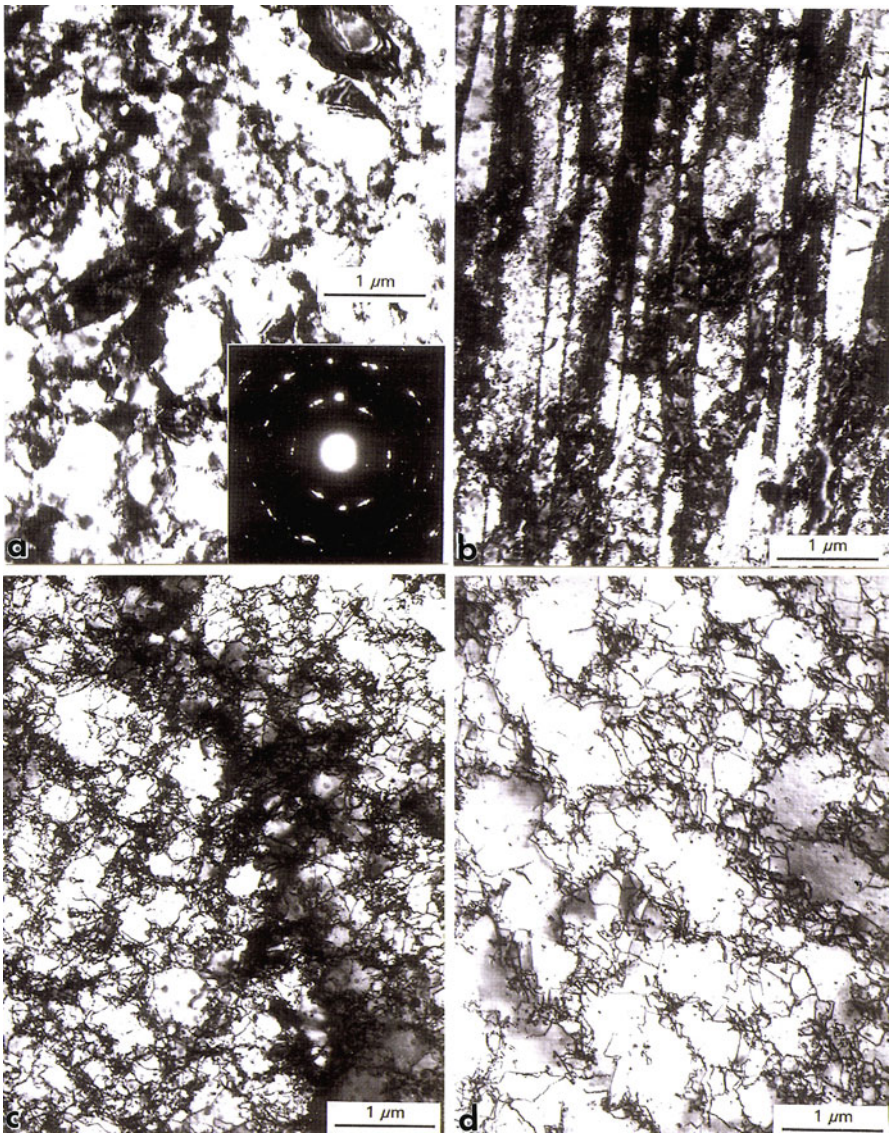


Fig. 20 TEM microstructure evolution in the Cu crater bottom region in Fig. 19a referenced to Fig. 13b. (**a**) DRX zone. SAED pattern insert shows random nanograin orientations. (**b**) Microbands ~3 mm from crater *bottom*. (**c**) Dislocation cells in PD2 zone below microband zone (PDI). (**d**) Dislocation cell structure corresponding to target base at ~15 mm from crater *bottom* (After Murr 2012)

Such model algorithms are embedded in so-called constitutive relationships or strength models.

To create the 3D and 2D simulations in Figs. 14 and 23, a Johnson–Cook (1983) constitutive relationship was used in the general form:

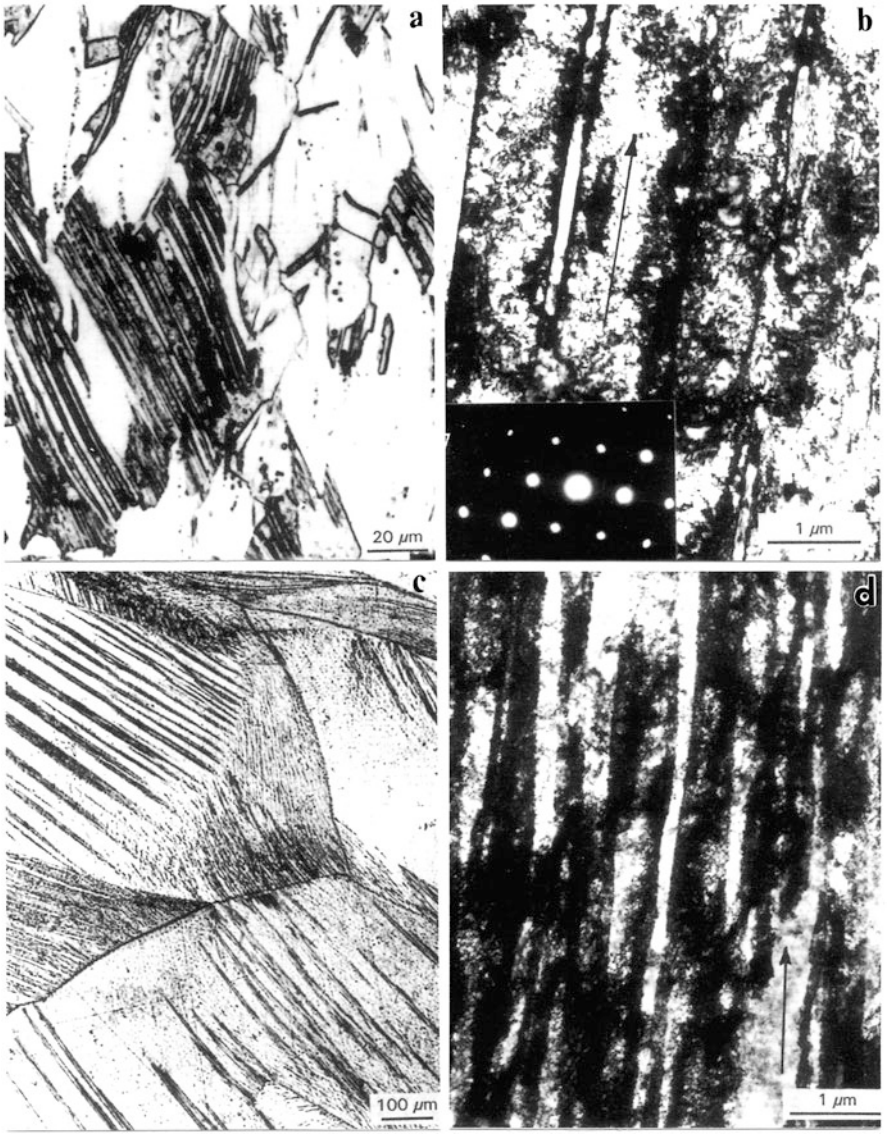


Fig. 21 Comparison of microband density (or volume fraction) and grain size for Cu impact craters represented by Fig. 13b in PD1 deformation zone. (**a**) Optical microscope image for 38 μm target grain size impacted by 3.2 mm diameter Al projectile at 6.01 km/s. (**b**) TEM image of microbands in (**a**). (**c**) Optical microscope image for 763 μm grain size target impacted by 3.2 mm diameter soda-lime glass projectile at 5.83 km/s. (**d**) TEM image of microbands in (**c**). *Arrows* in (**b**) and (**d**) are $[1\bar{1}2]$ directions in (110) fcc Cu grain surface as shown schematically in Fig. 18b (From Murr 2012)

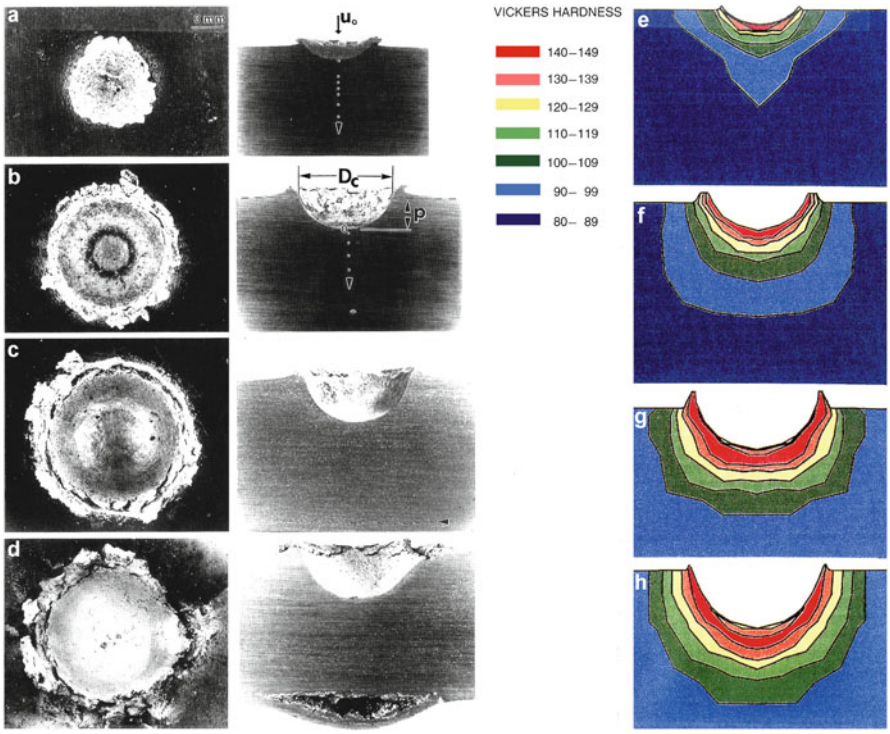


Fig. 22 Impact crater views for soda-lime glass (3.2 mm diameter) projectiles impacting Cu targets at (**a**) 1.08 km/s, (**b**) 2.64 km/s, (**c**) 4.4 km/s, and (**d**) 6.01 km/s. (**e**–**h**) show corresponding microindentation hardness maps corresponding to hardness zones indicated in color key. Note for Vickers hardness numbers: VHN 100 = 1 GPa. *Arrow* in (**c**) shows incipient spall crack (From Murr 2012)

$$\alpha_y = (\sigma_o + B\varepsilon^n)[1 + C\ln\dot{\varepsilon}/\varepsilon_o][1 - (T^*)^m], \tag{36}$$

where σ_o, B, C, n, and m are experimentally determined constants, ε is the strain, $\dot{\varepsilon}$ is the strain rate, $\dot{\varepsilon}_o$ is a reference strain rate (often equal to unity), and

$$T^* = \frac{(T - T_r)}{(T_m - T_r)}, \tag{37}$$

where T_m is the melting temperature and T_r is the reference temperature at which σ_r is measured and T is the temperature at which σ_y is calculated:

$$\sigma_y = \sigma_r[1 - (T - T_r)/(T_m - T_r)^m]. \tag{38}$$

To execute these computations, extensive libraries of data characterizing the material properties and behavior in the high-pressure state are utilized in computer programs or hydrocodes which can be adjusted through simple validations that can

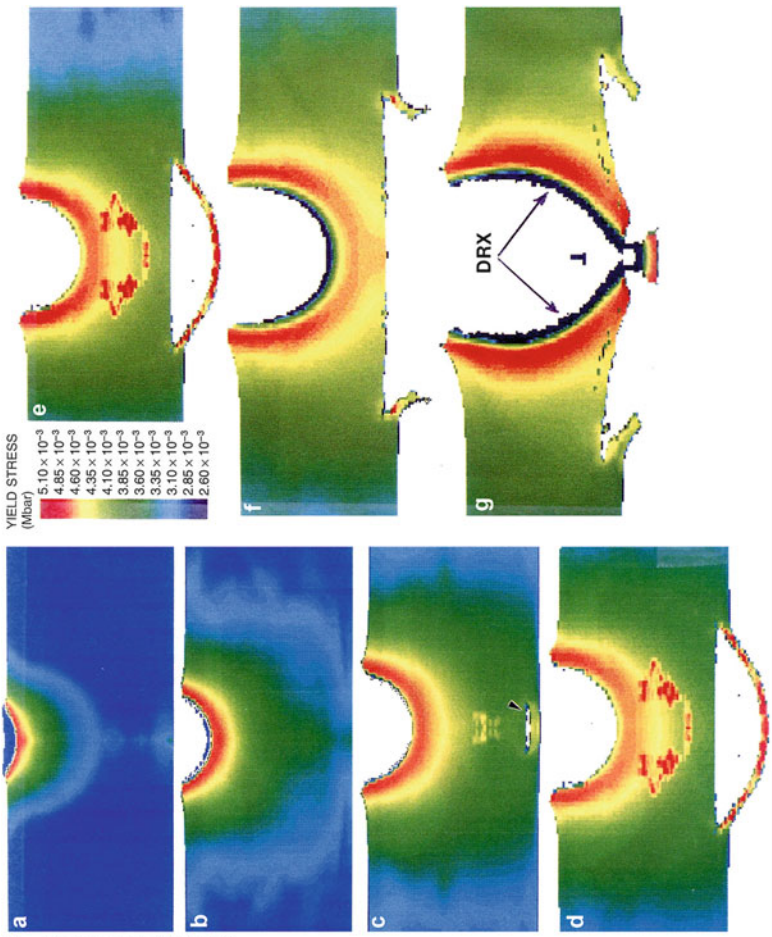


Fig. 23 2D computer simulations for impact craters in Fig. 22 showing residual yield stress maps (1 Mbar = 100 GPa): (**a**) $u_o = 1.08$ km/s, (**b**) $u_o = 2.64$ km/s, (**c**) $u_o = 4.4$ km/s, and (**d**) $u_o = 6.01$ km/s. *Arrow* in (**c**) shows incipient spall crack. (**e**–**g**) show impact at $u_o = 6.01$ km/s, 12 km/s, and 24 km/s, respectively. Color keys are inserted. Note ejected spall in (**f**) and (**g**) and penetration in (**g**). Copper target thickness is 1.3 cm (From Murr 2012)

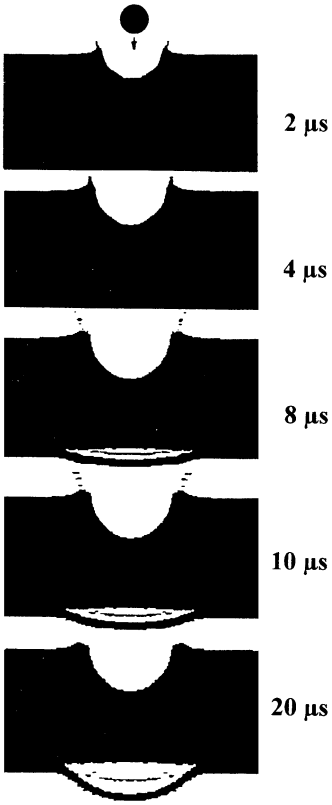


Fig. 24 2D computer simulations for crater development shown in Figs. 22d and 23d: 3.1 mm diameter Al projectile impacting Cu target at u_o = 6.01 km/s. Time sequence in μs. Note rim ejecta by DRX at 8 and 10 μs (From Murr 2012)

compare crater dimensions (ρ, D_c, h_r) shown in Fig. 13a for the final impact crater. Although this constitutes a single reference point, it can allow the simulation to be represented in time, as illustrated in Fig. 24 for the impact crater formation in Fig. 22d. More detailed features for hydrocode and related simulations and corresponding constitutive relationships will be discussed in Part XVI.

As noted earlier, there are anomalies in crater formation, especially for dense projectiles impacting soft targets, such as aluminum or aluminum alloys, although Eq. 22 provides some indication of penetration or crater depth as the projectile density to target density becomes very large. Figures 25 and 26 illustrate some of these features. In Fig. 25, a 3.2 mm diameter soda-lime glass penetrator (ρ = 2.3 g/cm^3) was impacted on an 1,100 Al target at u_o = 3.9 km/s. Figure 25a, b illustrates the resulting crater geometry, while Fig. 25c, d compares the residual hardness map and computed yield stress map surrounding the crater wall. The computer simulation in Fig. 25d shows a narrow DRX zone at the crater wall. In Fig. 25e, f, computer simulations extrapolate the yield stress maps to correspond to u_o = 5 km/s and 10 km/s, respectively, where the DRX zone at 10 km/s is more noticeable as indicated.

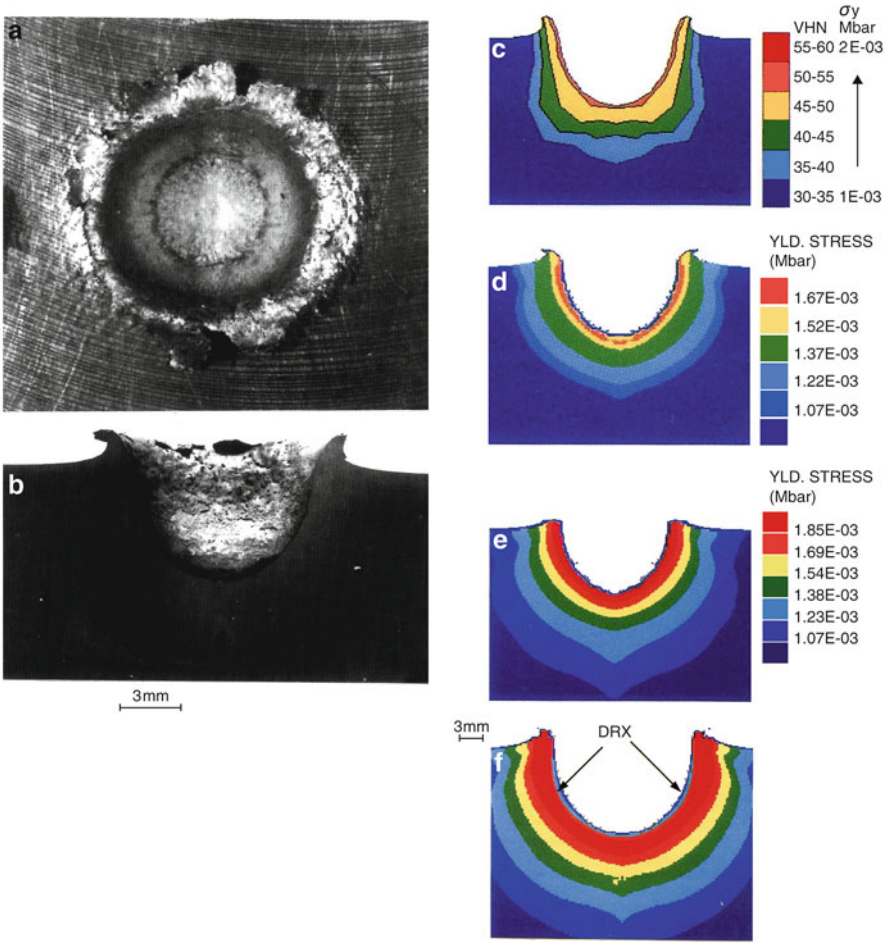


Fig. 25 Impact crater in 1,100 Al target formed by 3.2 mm diameter soda-lime glass projectile impacting at $u_o = 3.99$ km/s. (**a**) and (**b**) show surface and side-section views. (**c**) and (**d**) compare microindentation hardness and computed yield stress maps, respectively, according to color keys shown. 100 VHN = 1 GPa. 1 M bar = 100 GPa. (**e**) and (**f**) show extrapolated computer simulations at 5 km/s and 10 km/s, respectively. Note DRX zone in (**d–f**) (From Murr 2012)

In comparison to Figs. 25 and 26 shows the crater formation for a WC 3.2 mm diameter projectile ($\rho = 15$ g/cm^3) impacting the same 1,100 Al target ($\rho = 2.7$ g/cm^3) at $u_o = 2$ km/s. The crater depth, p, in Fig. 26a is nine times that of the crater depth in Fig. 25b, which exceeds the difference predicted by Eq. 22 by a factor of ~3. It can be observed in Fig. 26a that the WC projectile remains in the crater bottom, and consequently its full momentum excavated the long crater by forming a very narrow DRX zone at the crater wall which is almost imperceptible in the computer simulation in Fig. 26c. While the crater geometry is matched by

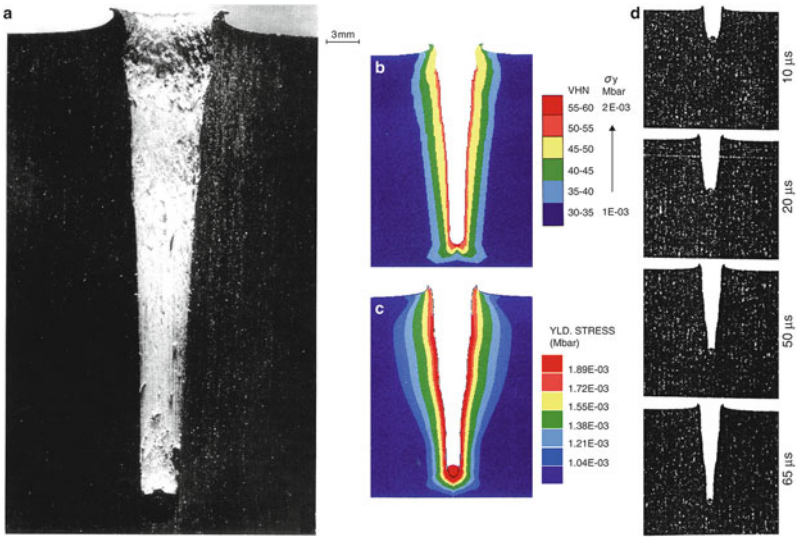


Fig. 26 Impact crater half-section views for 3.2 mm tungsten carbide (WC) projectile impacting 1,100 Al target at 2 km/s. (**a**) Experimental section view. (**b**) Microindentation hardness map for section in (**a**) and corresponding color key. (**c**) 2D section computer simulation of residual yield stress map of (**a**) and corresponding color key. 100 VHN = 1 GPa; 1 Mbar = 100 GPa. (**d**) Computer simulation of time sequence development for impact crater in (**a**). Time in μs (From Murr 2012)

the simulation in Fig. 26c, there is a notable difference in the residual hardness and yield stress profile maps on comparing Fig. 26b, c. Figure 26d illustrates the simulated crater formation over a time frame of 55 μs.

The significance of Fig. 26 is that for thin Whipple or multilayer shields on space vehicles (Christiansen and Kerr 2001), which provide protection against hypervelocity penetration of low-density penetrators, slow-moving, high-density space debris could have deleterious effects because there may not be any projectile fragmentation, and penetration could be correspondingly significant (Valerio-Flores et al. 2004).

Rod Penetration and Target Plug Formation

When the penetrator geometry changes from a sphere to an extended rod, the shock wave propagation into the projectile does not allow for fragmentation, melt, or vaporization at ballistic velocities which do not exceed 3 km/s. Projectile momentum becomes a significant contributor to penetration, similar in some respects to dense projectiles, such as the WC projectile forming an extended crater in Fig. 26a. Nonetheless as the target is penetrated by continuous excavation of the extended crater, crater material flows along the penetrator to the entrance surface where it is ejected through tensile jetting of the rim material. This solid-state target flow occurs

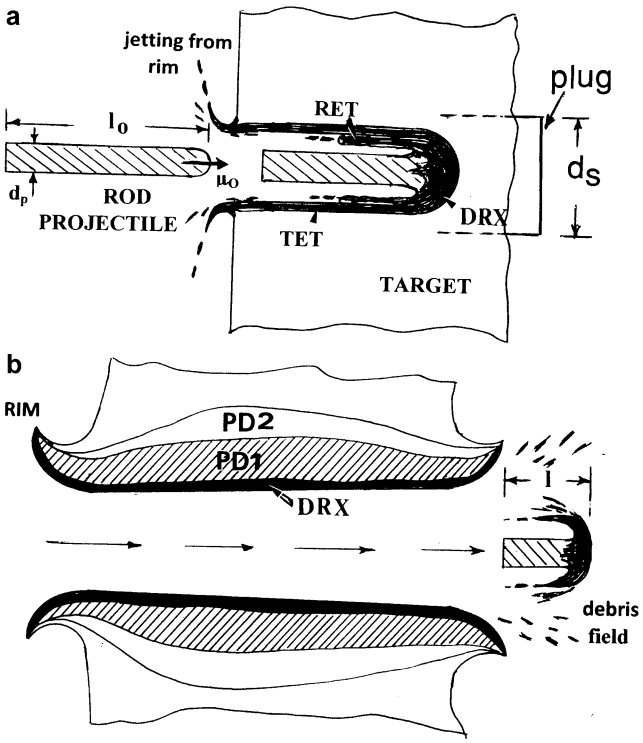


Fig. 27 Long-rod (normal) penetration into a finite target of thickness z_o at initial velocity u_o. (**a**) Penetrating and eroding rod and target by DRX forming a DRX cylinder or tube: rod erosion tube (RET) and target erosion tube (TET). Plug or target dimension after projectile penetration is denoted z. Target velocity v_o at $t = 0$; v at $t > 0$. (**b**) Target perforation showing DRX and plastic deformation zones (PD1 and PD2) as in Fig. 13b. Residual projectile emerging from perforated target has a length ℓ_r

by dynamic recrystallization (DRX), and the flowing target DRX material forms a target erosion tube next to the penetrating rod. In addition, the rod also erodes by dynamic recrystallization, and the eroding rod DRX material also forms an erosion tube next to the target erosion tube (TET). Material in the rod erosion tube (RET) also flows backwards, in the direction of excavated target material in the TET, and is also ejected at the rim. These features are illustrated schematically in Fig. 27. In some cases the DRX material in the RET and TET can flow together to form mechanically alloyed zones or other thermodynamic phase regimes if local temperature elevations can support such reactions.

Grace (1995) initially developed the mechanics for rod penetration into finite thickness targets represented in Fig. 27a. The penetration process considered deceleration of the rod mass and acceleration of target mass as well as the simultaneous erosion of each material body as shown for DRX formation of the rod erosion tube (RET) and the target erosion tube (TET) shown schematically in

Fig. 27a. Later and more recent analysis by Grace (2001) illustrated that the rod's kinetic energy is distributed between the kinetic energy of the residual rod, the target plug, and the corresponding erosion products (RET and TET in Fig. 27a), as well as the work related to deformation of the target and during the penetration process.

Following these analytical developments (Grace 1995, 2001), the respective equations of motion for the rod (projectile) and target (Fig. 27a) are

$$\left.\begin{aligned} M_p \frac{d(u - v_o)}{dt} &= -S_p A_p \\ \frac{d\ell}{dt} &= -(u - v_o)\,\text{rod erosion} \\ M_p &= \rho_p A_p \ell \end{aligned}\right\} \tag{39}$$

and

$$\left.\begin{aligned} M_t \frac{dv}{dt} &= -S_t A_t \\ \frac{dz}{dt} &= -v\,(\text{target erosion}) \\ M_t &= \rho_t A_t z \end{aligned}\right\} \tag{40}$$

where M_p and M_t are the projectile and target mass, S_p and S_t are the nominal projectile and target strengths (yield strengths), A_p and A_t are the projectile and target cross-sectional areas, ρ_p and ρ_t are the projectile and target densities, and v is the penetration rate. The target thickness is z and z_o. z in Fig. 27a represents the distance from the head of the projectile in the target to the back side of the target (or the plug surface). From Eqs. 39 and 40, the corresponding equations of motion for the projectile and target are obtained:

$$\text{Projectile} : (u - v_o) = (u_o - v_o)\left[1 + \frac{2S_p}{\rho_p (u_o - v_o)^2} \ln\left(\frac{\ell}{\ell_o}\right)\right]^{1/2} \tag{41}$$

$$\text{Target} : v = v_o \left[1 + \frac{2S_t}{\rho_t v_0^2 2} \ln\left(\frac{z}{z_o}\right)\right]^{1/2} \tag{42}$$

If the penetrating rod perforates the target (Fig. 27b), its residual velocity, u_r, is given by

$$U_r \left(\frac{M_r V_c 2 - M_c v_0^2 - 2E_f}{M_r + M_c}\right)^{1/2} \tag{43}$$

where E_f is the energy required for plug separation:

$$E_f = \pi d_s S_s z_c{}^2. \tag{44}$$

The plug diameter, d_s, is shown in Fig. 27a, while S_s in Eq. 44 is the target shear stress, and z_c is the plug length. The other parameters in Eq. 43 are

$$V_c = \left(\frac{2\,S_t}{\rho_p}\right)^{1/2} \tag{45}$$

$$M_r = \pi \frac{\left(\rho_p {d_p}^2 \ell_r\right)}{4} \tag{46}$$

$$M_c = \pi \frac{\left(\rho_t {d_s}^2 Z_c\right)}{4} \tag{47}$$

The residual rod length, ℓ_r, in Eq. 46 is illustrated in Fig. 27b. It is implicit in Fig. 27b that a plug has not formed and the target is perforated, producing the residual (eroded), penetrated projectile. In this case, M_c and E_f are zero. Alternatively, the rod projectile can come to rest in the target, with some residual rod length remaining in the penetration cavity; or the rod can be completely consumed or eroded within the penetration cavity. For the ideal, hydrodynamic case, the penetration depth will be given by Eq. 22. In the penetration process, the kinetic energy lost by the rod projectile appears in rod erosion products (RET in Fig. 27a), flow of target material forming the TET and rim, as well as ejected erosion (DRX) material at the rim. In addition, strain energy to deform the rod and target materials during radial expansion, target mass motion, and heat also contributes to the penetration process.

Figure 28 shows details and computer simulations for a tungsten-alloy (W-7 Fe, Ni, or tungsten heavy alloy: WHA) rod having a length/diameter ratio (ℓ_o/d_p) of 10, penetrating (and perforating) a 3.5 cm thick Cu target at an initial velocity $u_o =$ 1.5 km/s. Figure 28a shows a flash X-ray image of the eroded rod emerging from the target along with a debris field formed as the projectile perforates the target. In this penetration condition, there is no plug formation, but at some distance from the back surface of the target, there is some reversal of the DRX target flow and the TET and RET particulate into an emerging debris field. These features are observed in the computer simulations in Fig. 28c, d. In Fig. 28c, the penetration cavity geometry shown in the half section of Fig. 28b was used to validate the simulation. In addition, ℓ/ℓ_o in Eq. 41 can be determined from the emerging projectile length (ℓ) in the flash X-ray image in Fig. 28a. Figure 28d shows a time sequence illustrating the rod penetration and perforation as well as the debris field formation at the back of the penetration channel cavity. Figure 28b also illustrates the remnant DRX flow zone or TET in the penetration cavity wall, which also contains some RET (WHA) mixing. Plastic deformation zones implicit in Fig. 27b are also shown in Fig. 28b. This feature is also illustrated in the measurement of residual microindentation hardness in the half section of Fig. 28b, as shown in Fig. 29b.

The circled rim region in the penetrated Cu target of Fig. 28b is shown enlarged in the optical microscope image of Fig. 30. The DRX zone, composed of layered

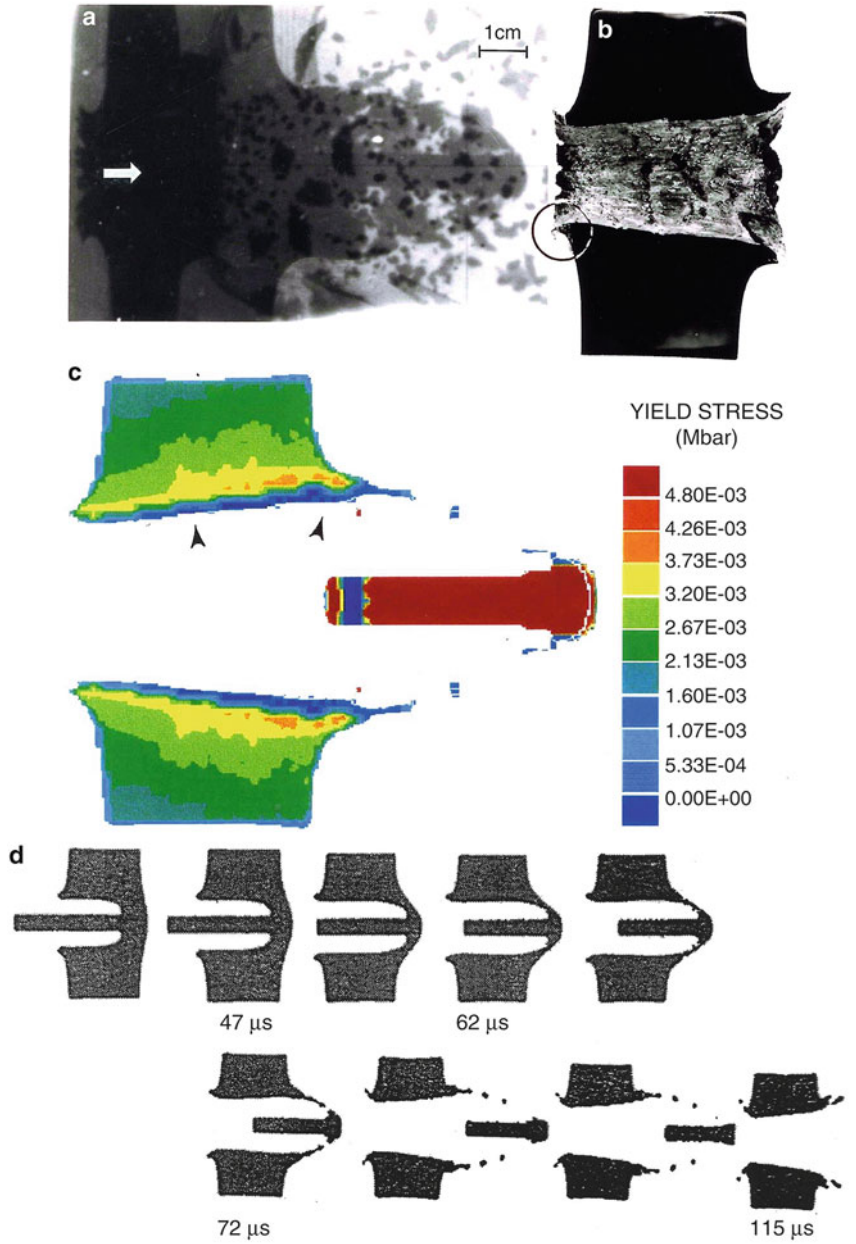


Fig. 28 Tungsten heavy alloy (WHA) rod ($d_p = 0.78$ cm, $\rho_p = 18$ g/cm^3) perforating 3.5 cm thick Cu target ($u_o = 1.5$ km/s). (**a**) Flash X-ray image showing eroded rod and debris field emerging from back side of target. (**b**) Target half section after perforation. (**c**) 2D computer simulation showing residual yield stress map (color key: 1 M bar = 100 GPa). (**d**) 2D computer simulation showing penetration time sequence (in μs) (From Murr 2012)

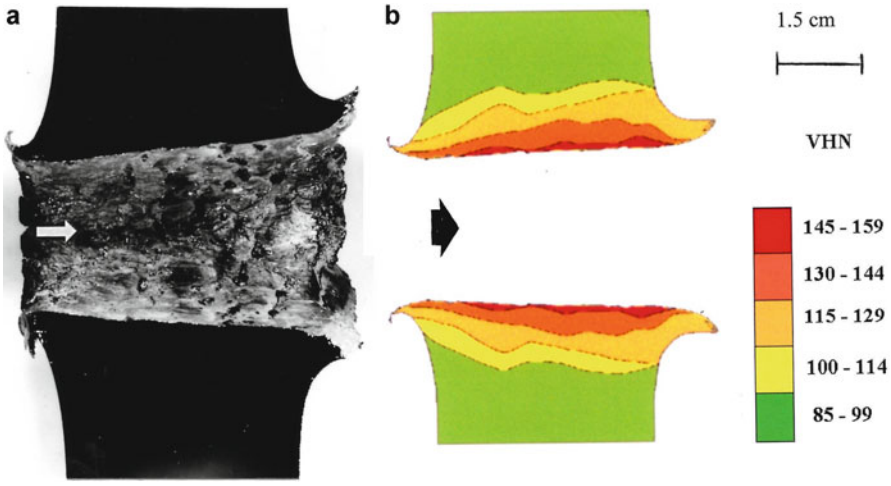


Fig. 29 WHA rod-perforated Cu target half section as in Fig. 28a, b. (**c**) Perforated Cu target half section. (**b**) Microindentation hardness map corresponding to the half-section ion (**a**). Color code: 100 VHN = 1 GPa (After Murr 2012)

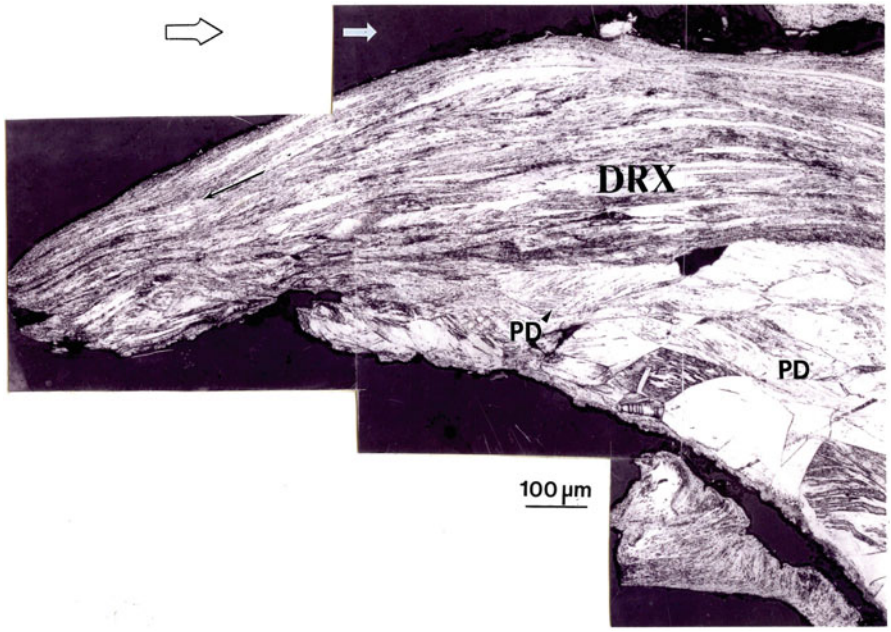


Fig. 30 Optical microscope image showing DRX microstructure for perforated Cu target rim section shown *circled* in Fig. 28b. *Arrows* pointing *right* show projectile (rod) penetration direction. *Black arrow* at *left* shows DRX microstructure flow. *Top* surface is inside cavity surface. Plastic deformation of target grains is shown at PD

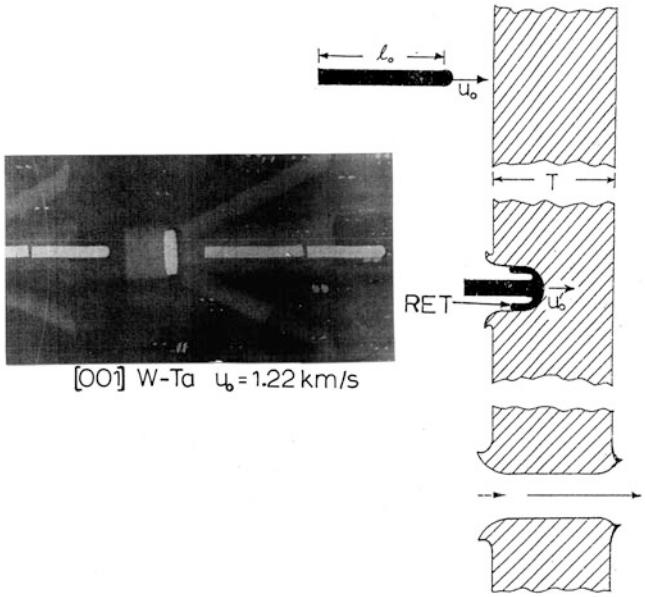


Fig. 31 Flash X-ray image of W-4 % Ta [001] single-crystal rod projectile launched from a sabot system (S) using a pusher plate (PP). Schematic views show the initial rod impacting at u_o velocity and penetrating at a velocity $u_o' < u_o$ as DRX-forming RET is created by projectile erosion

flow regimes and measuring ~200 μm thick, is shown in relation to the plastically deformed (PD) zone containing complex microband structures in the variously distorted Cu target grains.

Ballistic penetrators as long rods have been used for decades as effective armor piercing rounds fired from 120 mm bore or similar guns, with aspect ratios (ℓ_o/d_p) of 15. These rounds are traditionally WHA compositions, but a great deal of research continues to explore W (as single-crystal orientations, amorphous structure, etc.) as well as other alloys such as W-4 Ta, W-10 Ag, Ta-10 Ag, etc. For single-crystal W or W alloys, [001] orientations recrystallize and flow by DRX more efficiently than other orientations, and efficient flow (and resulting target penetration) has also been observed in amorphous or very small nanograin structures. Figure 31 illustrates a W-Ta penetrator rod fired from a gun using a so-called pusher plate and a sabot system observed by flash X-radiography. Related penetration and perforation schematics are also shown. Figure 31 illustrates cleavage of the [001] W-Ta rod as a consequence of a small velocity gradient along the rod length. The penetration schematics shown in Fig. 31 are expanded somewhat in Fig. 32 which illustrates the formation of a target plug and its expulsion from the penetrated target cavity.

Figure 33 shows a cross-sectional view of a W [001] single-crystal rod projectile penetrating a rolled homogeneous armor (steel) (RHA) target where the rod is almost completely eroded by DRX structure formation and flow forming an RET of fragmented DRX as shown by arrows. A region of target erosion (TET) is also noted along the penetration channel walls. Figure 34a shows deformation twins

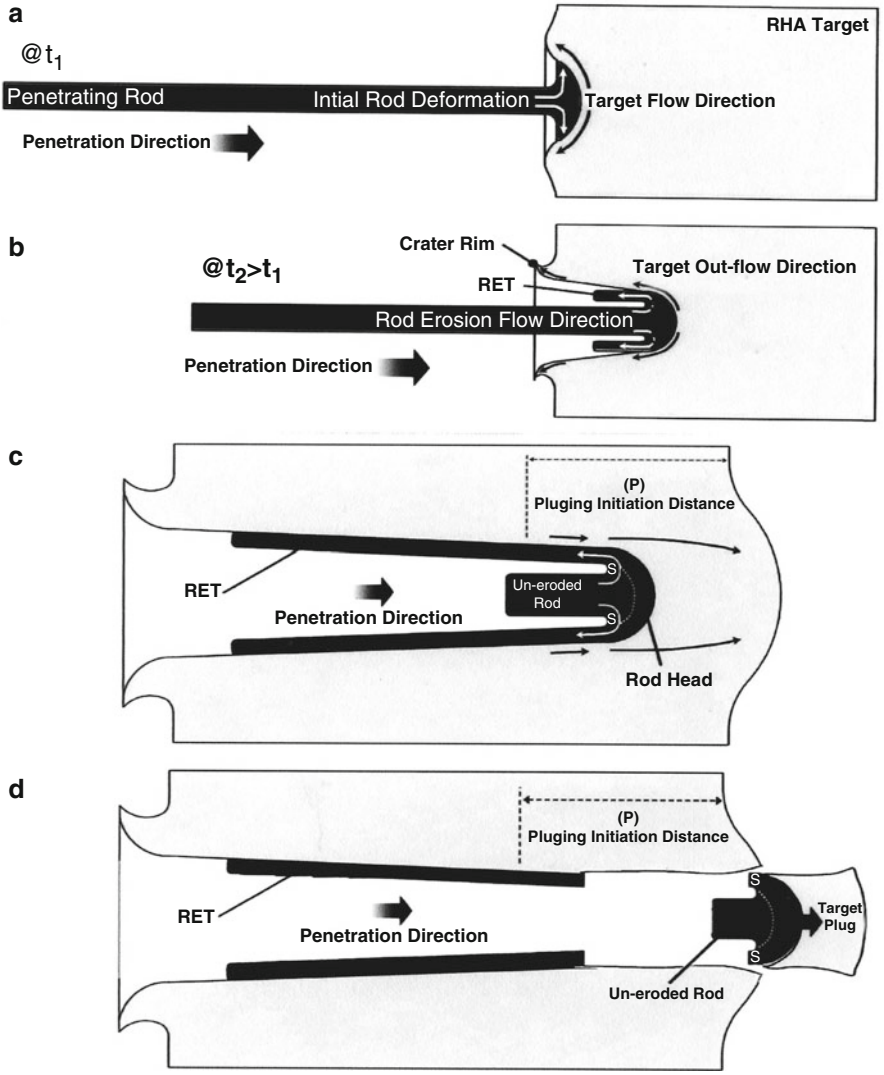


Fig. 32 Schematic details for projectile rod impacting and interacting with rolled homogeneous alloy (steel) (RHA) target (Courtesy of Dr. Carlos Pizana)

(T) near the rear of a W-4 Ta [001] projectile which has not been eroded to the extent of Fig. 33, while Fig. 34b shows complex vortex formation similar to the RET DRX flow regimes illustrated in Fig. 33. Figure 35a compares the [001] single-crystal dislocation structure in a W precursor rod observed by TEM in an electron-transparent slice cut perpendicular to the [001] rod axis, with the RET DRX (nano) grain structure (similar to the insert in Fig. 34b) in Fig. 35b. Deformation twins have also been observed in W rod ballistic penetrators (Pappu et al. 2001).

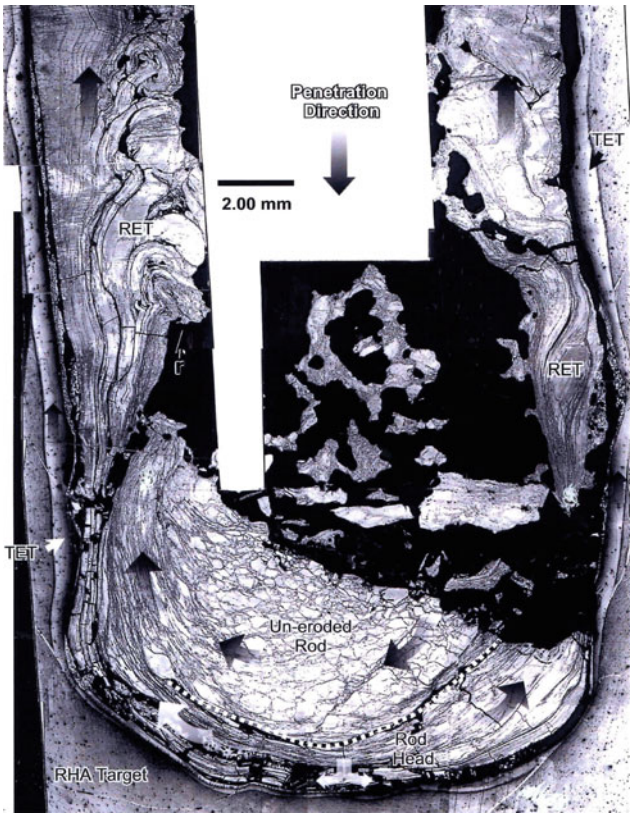


Fig. 33 Optical (light) micrograph composite showing a residual, uneroded [001] single-crystal W rod head in an RHA steel target cavity after penetrating at an initial impact velocity of 1.35 km/s. RET and TET refer to rod and target erosion tubes, respectively (From Murr and Pizana 2007)

It might be noted in retrospect that adiabatic shear bands (SB) are also formed within the DRX structures and in fact compose localized shear zones in the single crystal as illustrated in Fig. 34a. These bands also initiate fracture forming single-crystal blocks as part of the rod deformation, and they facilitate the detachment and ejection of a target plug as shown for a W [001] penetrating rod in an RHA target in Fig. 36. It can be noted that adiabatic shear bands (SB) formed as shown in Fig. 34a and along the plug and associated cracking as shown in Fig. 36 are roughly an order of magnitude smaller in width than those DRX flow regimes associated with crater flow and rim formation as shown in Fig. 30.

It can be observed that for large, blunt projectiles having a diameter approximating the target thickness and having velocities in the ballistic range (1–2 km/s), the projectile will not melt according to Eq. 33, for example. Consequently, at some critical impact velocity, u_o corresponding to $d_p/T \cong 1$, where T is the target thickness, a target plug will be expelled from the target, allowing complete target

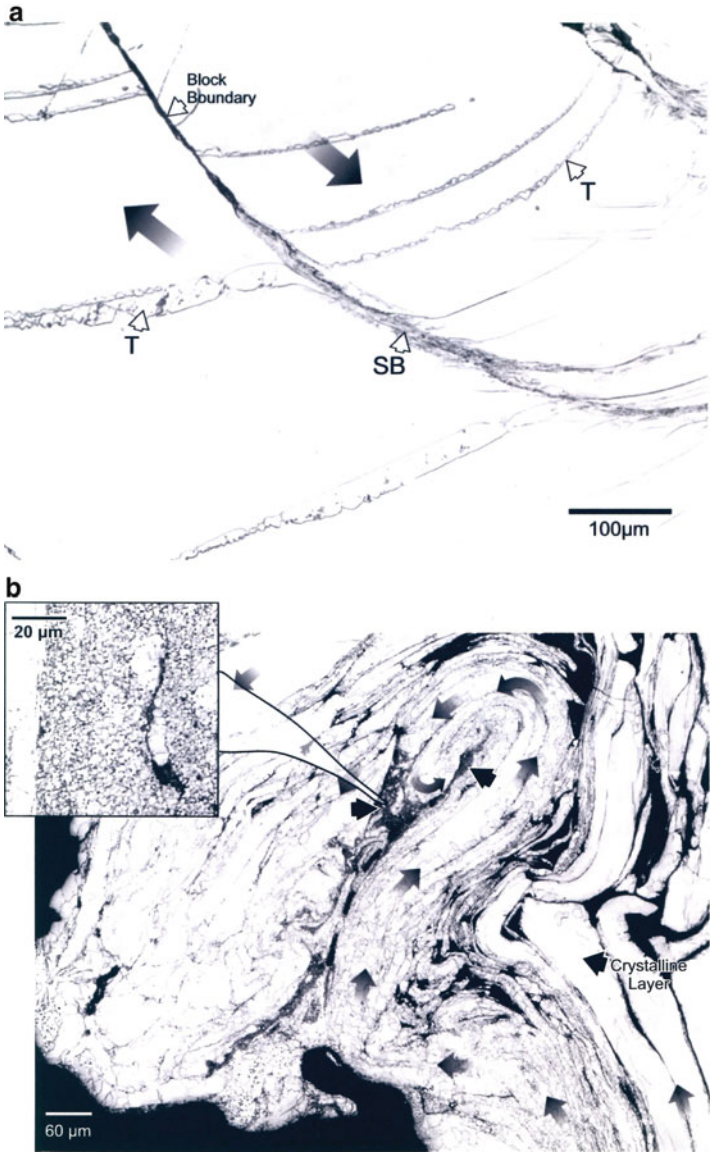


Fig. 34 Deformation in W-4 Ta [001] single-crystal penetrator rods in RHA target. (**a**) Deformation twins (T) and shear bands (SB) and crack nucleation (block boundary). *Arrows* illustrate shear flow. (**b**) Complex DRX vortex flow regime associated with projectile RET formation similar to Fig. 33. *Arrows* indicate solid-state flow regimes (Courtesy of Dr. Carlos Pizana)

perforation. Figure 37 illustrates this plugging phenomenon for 4,340 steel projectiles impacting Ti-6Al-4 V targets at velocities shown in the sequence Fig. 37b–d. Figure 37e illustrates the incipient plug formation and expulsion with

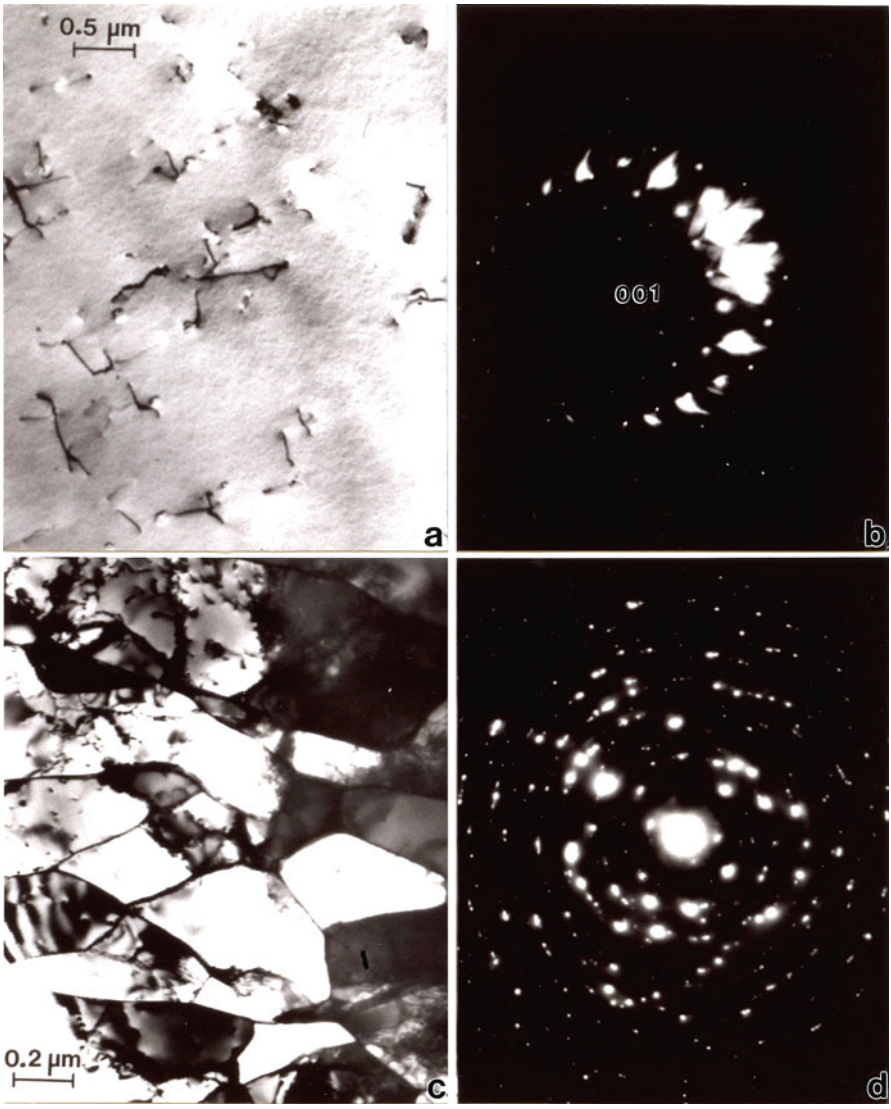


Fig. 35 TEM images and corresponding selected-area electron diffraction (SAED) patterns for initial [001] W rod projectile single-crystal (**a**) and residual DRX erosion product in an RHA target (**b**)

reference to the plugging schematic shown in Fig. 37a, and the actual plug formation and expulsion shown in the sequence of Fig. 37b–d. The adiabatic shear bands (ASBs) form normal to the impact axis and along the plug flow zone. In fact the ASB formation from localized DRX zones, and subsequent crack formation within these zones, allows the target plug to be expelled by the projectile. Figure 38a, b

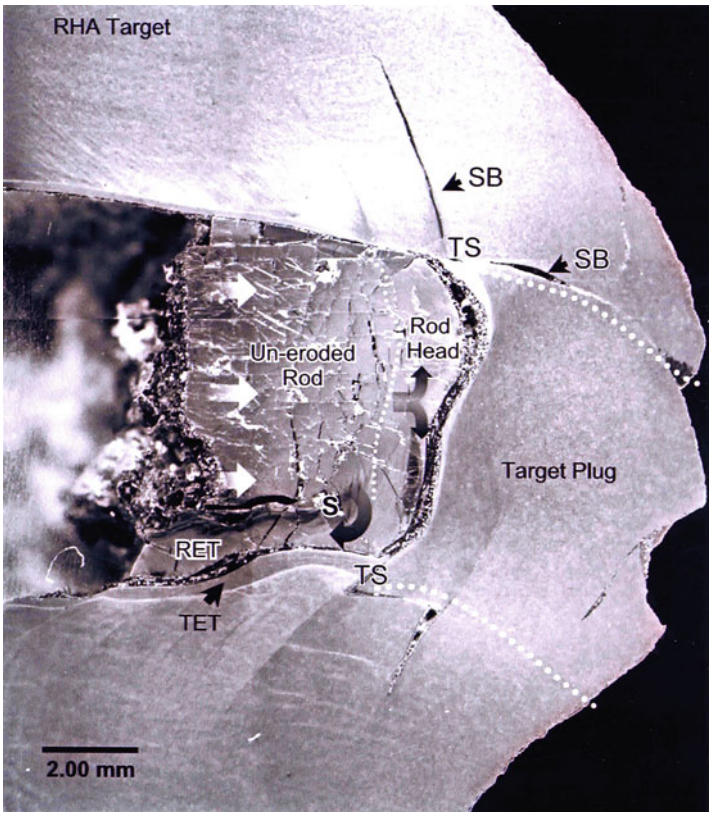


Fig. 36 [001] W penetrator rod clad with 140 μm thick Inconel 718 ($u_o = 1.33$ km/s) penetrating an RHA target. RET and TET initiation at the rod head are denoted S and TS. Shear band and shear band crack nucleation are denoted SB (From Murr and Pizana 2007)

illustrates examples of ASB formation and cracking, which is characterized by crack nucleation and coalescence within the shear bands. The shear band thickness (Δ_s) in Figs. 37e and 38 can be expressed by Dodd and Bai (1989)

$$\Delta_s = 2\left(\frac{\lambda\Theta}{\tau\dot{\gamma}}\right)^{1/2} \tag{48}$$

where

$$\Theta = 0.9\tau\dot{\gamma}/\rho \cdot C; \tag{49}$$

τ is the shear stress, $\dot{\gamma}$ is the shear strain rate, ρ is the target material density, and C is the heat capacity. λ in Eq. 48 is the corresponding coefficient of heat conduction. Similarly, the localized shear strain, γ, associated with the formation of the DRX–ASB can be expressed by

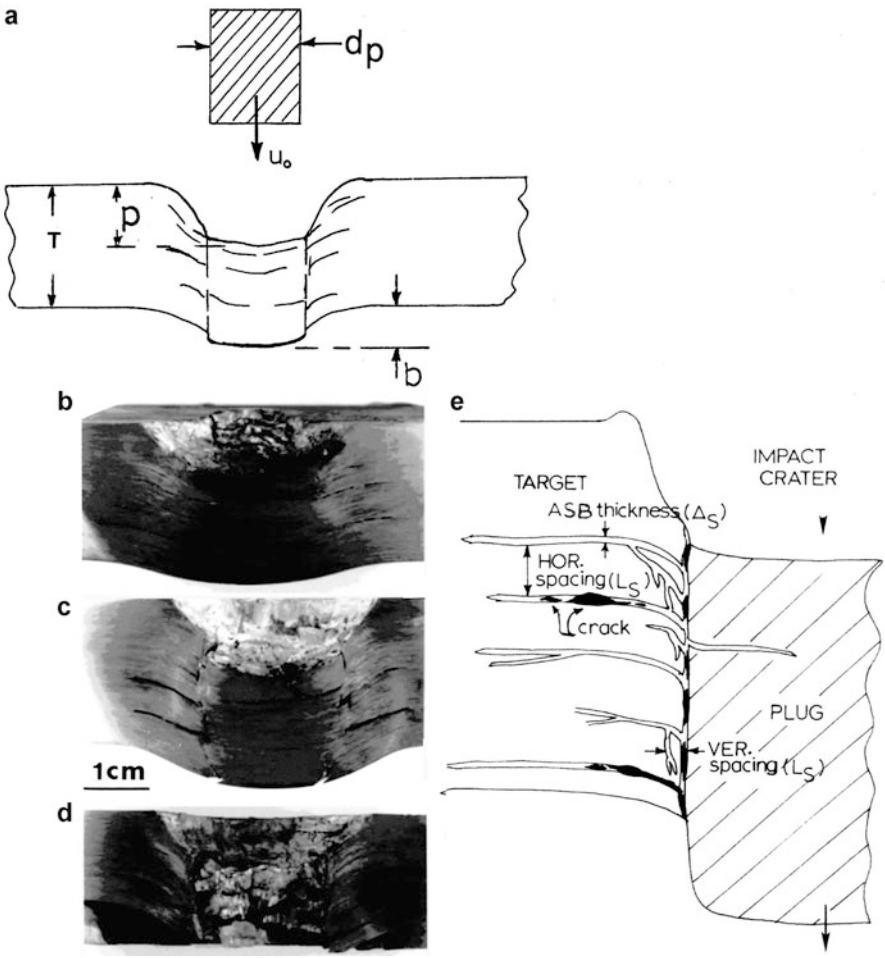


Fig. 37 Ballistic plug formation and associated adiabatic shear band (ASB) generation and cracking. (**a**) Impact of blunt projectile ($d_p \sim T$) forming target plug which emerges a distance b on back target surface (schematic). (**b–d**) show Ti-6Al-4 V targets impacted by 4,340 steel projectiles ($d_p = 2$ cm) at velocities of (**b**) 633 m/s, (**c**) 1,006 m/s, and (**d**) 1,027 m/s. The ballistic limit was $u_o \cong 2{,}016$ km/s. (**e**) Horizontal (HOR) and vertical (VER). ASB formation associated with ballistic plugging. ASB width is denoted Δ_s. ASB spacing is denoted L_s. Schematic corresponds to half-section schematic in (**a**) (Adapted from Murr et al. 2009)

$$\gamma = \frac{\dot{\gamma}\rho C \Delta_s 2}{0.36\lambda}. \tag{50}$$

Since $\dot{\gamma}$ is related to the impact velocity (increasing with increasing u_o in Fig. 37a), the associated shear strain characterizing the ASBs can be illustrated somewhat schematically in ASB formation and crack nucleation as shown in Fig. 39, where it can be noted that dislocation density increases as the DRX zone

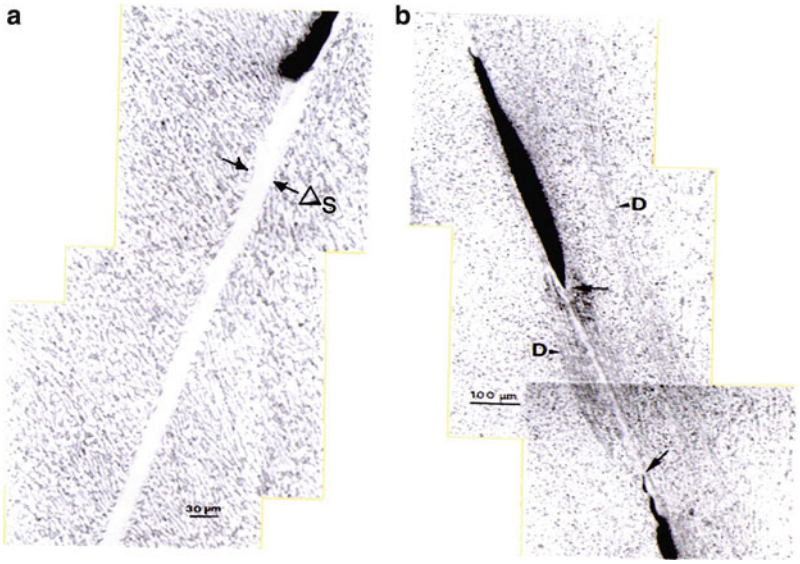


Fig. 38 Adiabatic shear bands in impacted Ti-6Al-4 V targets. (**a**) Blunt 4,340 steel projectiles impacting at $u_o = 794$ m/s. (**b**) Blunt 4,340 steel projectiles impacting at $u_o = 905$ m/s. *Arrow* shows blunting crack. D shows deformation zones which are precursors to and associated with ASB. *Arrows* show advancing and nucleating cracks

forms, and this work hardens the ASB, causing the fracture stress to be exceeded and nucleating cracks within the ASB. Murr et al. (2009) have in fact measured a roughly 17 % increase in hardness across ASBs as shown in Fig. 38 in contrast to the base target hardness.

Shaped Charges, Explosively Formed Penetrators, and Railguns

In 1888, a chemist named Charles Munroe working at the US Naval Torpedo Station in Newport, Rhode Island, noticed that when a block of explosive guncotton with the manufacturer's name stamped onto it was detonated next to a metal plate, the lettering in the name was cut into the plate. Munroe found that raising the lettering on the explosive also raised the lettering on the metal plate, and by 1894, he constructed what has become known as a shaped charge by arranging sticks of dynamite around a can placed on a few inches of iron and steel plates, which upon detonation created a three inch diameter hole through the plates. During WWII, this concept of a hollow charge lined with a metal became a military warhead or shaped charge warhead which was especially effective in antitank warfare. Contemporary shaped charge warheads using conical metal liners surrounded by a high explosive are referred to as high-explosive antitank (HEAT) warheads which are configured in guided missiles, rockets, gunfired projectiles, and other weapons configurations.

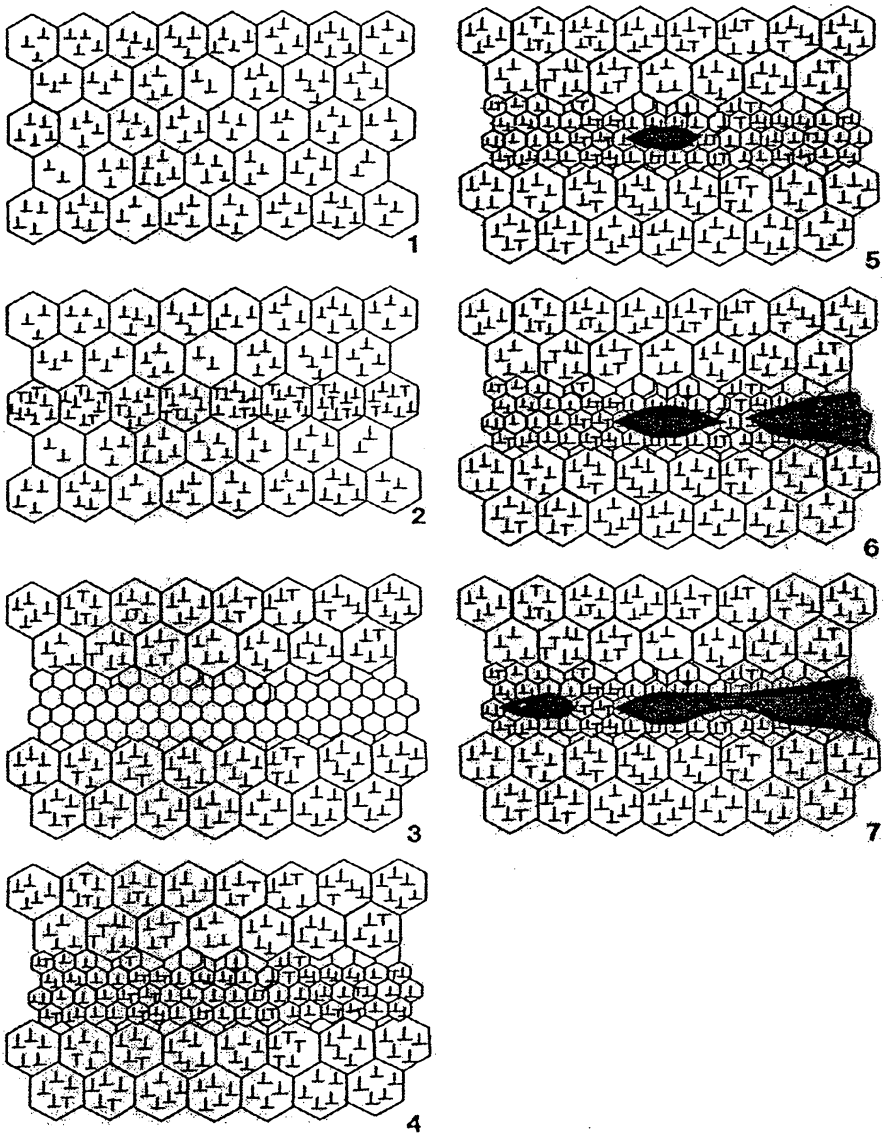


Fig. 39 Schematic illustration of adiabatic shear band microstructure evolution. (*1*) Deformed, equiaxed target grain structure with dislocations denoted by ⊥. (*2*) Shear localization creating dense dislocation zone. (*3*) Incipient ASB by DRX. (*4*) Deformation of DRX in ASB (forming dislocations in DRX grains). (*5*) Crack nucleation in deformed ASB–DRX grains. (*6*) Crack growth and (*7*) nucleation (From Murr et al. 2009)

Figure 40a, b illustrates the fundamental features for a conical shaped charge (a high explosive (HE) surrounding a conical metal liner). The detonating HE with a detonation front traveling at a velocity V_D (detonation velocity shown for different

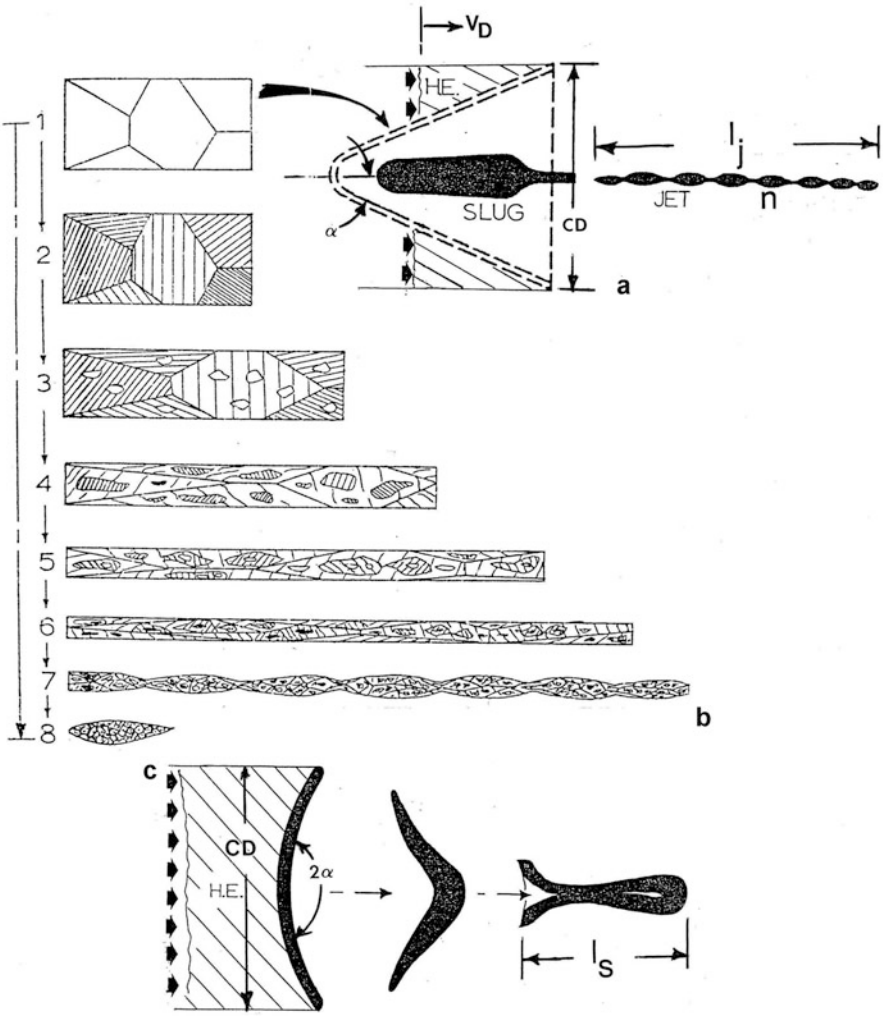


Fig. 40 Shaped charge schematics. (**a**) Conical liner shaped charge which upon detonation forms a slug and a hypervelocity jet. The charge or cone diameter is denoted DC. ℓ_j is the jet length which can particulate into n jet particles. The apex half angle is α. V_D is the detonation velocity at the detonation front in the exploding high-explosive charge (HE). (**b**) shows a series of schematic views for the formation of the jet by DRX (*1–8*). The equiaxed grain structure at (*1*) characterizes the liner cone. The liner grain structure deforms, forming DRX grains which can continue to deform and form new DRX grains which can also grow. (**c**) shows an explosively formed penetrator (EFP) or self-formed penetrator for a liner having a large angle α (or $2\alpha \gtrsim 120°$)

explosive compositions in Table 1) collapses the cone starting at the apex, where the line material of mass, m, flows by DRX in the solid state, to the right and left (outside and inside liner surfaces) of the apex to form a slow-moving slug and fast-moving het of liner metal having an initial diameter of d_j as shown schematically in

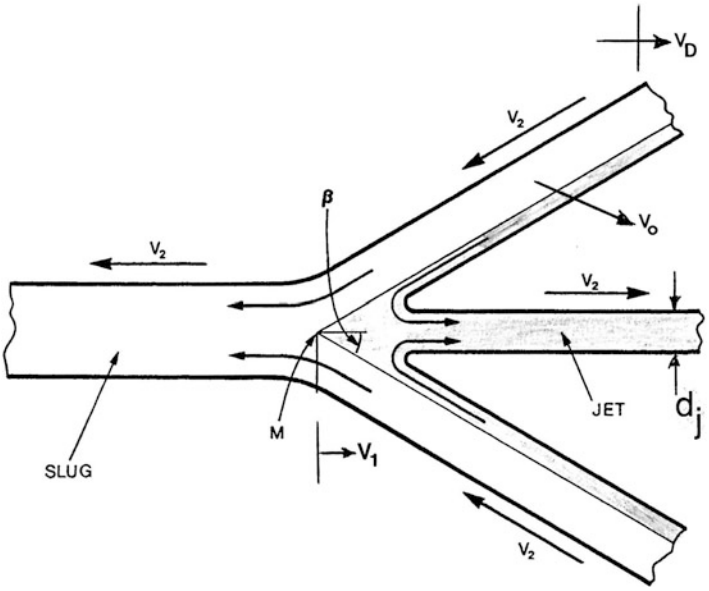


Fig. 41 Schematic showing detonation-initiated cone collapse in a shaped charge as in Fig. 40a with jet and slug formation by liner solid-state (DRX) flow. The collapse angle is denoted β. At the onset of collapse, $\alpha = \beta$. The liner is moving at a velocity V_o which moves the stagnation point at M at a velocity V_1. Liner flow occurs at a velocity V_2. The detonation velocity of the explosive is denoted V_D. The initial jet diameter is d_j

Fig. 41. Considered in the context of a stationary frame of reference, the original apex point at M, called the stagnation point in the collapsing cone system, moves at a velocity V_1. Correspondingly, the jet and slug move at velocities:

$$\begin{aligned} V_j &= V_1 + V_2 \\ V_s &= V_1 - V_2 \end{aligned} \tag{51}$$

As the cone collapses, the original apex angle or half angle, α, shown in Fig. 40 becomes the collapse angle, β, in Fig. 41. These angles are related by

$$\alpha = \cos^{-1}\left[\left(\frac{2V_D}{V_o}\right)\sin\left(\frac{\beta-\alpha}{2}\right)\right]. \tag{52}$$

Similarly, the liner mass, m, is given by

$$m = m_j + m_s, \tag{53}$$

where the jet mass, m_j, and the slug mass, m_s, are given by

$$\begin{aligned} m_j &= m\,(1-\cos\beta)/2 \\ m_s &= m\,(1+\cos\beta)/2. \end{aligned} \tag{54}$$

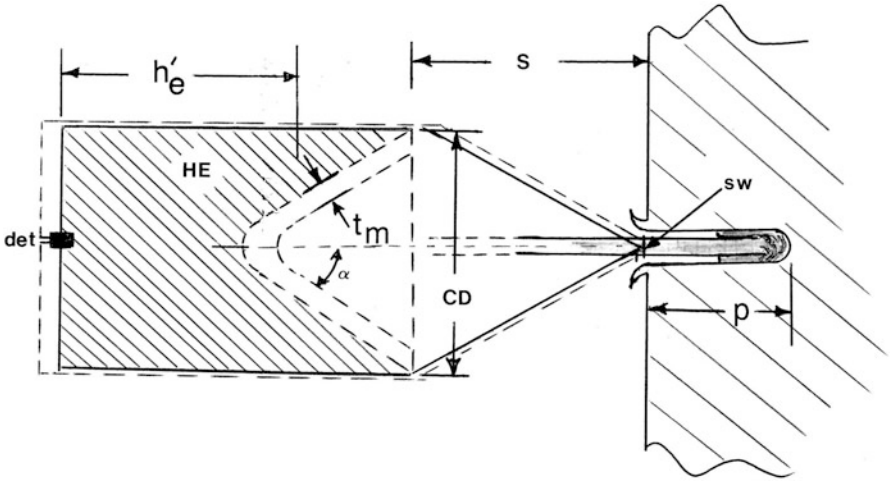


Fig. 42 Conical shaped charge detonation geometry and target penetration (P) schematic liner with thickness t_m and mass M is backed by high explosive (HE) having a mass C and mean effective thickness h_e contained in a shell housing having a conical front extension having a designed height for a desired standoff, S. On contact with the target, a simple "doorbell" switch (SW) activates the detonator (det) which initiates the detonation of the HE. The penetration, P, is often designated in equivalent charge or cone diameters, CD, or N (CD); $N = 1, 2$, etc., or a p/CD ratio. Standoff, S, is also expressed in cone diameters: S = N (CD)

The collapse velocity, V_o, which is assumed to be constant over the cone collapse event (Fig. 41), is related to the detonation velocity, V_D (Table 1), by

$$V_o \frac{2V_D \sin(\beta - \alpha)/2}{\cos \alpha}. \tag{55}$$

This velocity can also be approximated from the plane, open-sandwich arrangement for plane shock wave production in Fig. 1a:

$$V_d \cong \sqrt{2E}\left[\frac{3}{1 + 5(M/C) + 4(M/C)^2}\right]^{1/2} \tag{55a}$$

where $\sqrt{2E}$ is the *Gurney velocity* for a particular explosive composition as shown in Table 1, and

$$(M/C) = \rho_d h_d/\rho_e h_e = \rho_m t_m/\rho_e h'_e \tag{56}$$

as shown in Fig. 42, where t_m is the liner thickness and h'_e is the average or effective explosive height relative to the cone apex. Of course it is more accurate to simply consider the mass ratio, M/C. It is also observed in Fig. 41 that

$$V_1 = V_o \cos((\beta - \alpha)/2)/\sin\beta, \tag{57}$$

and

$$V_2 = V_o\left[\frac{\cos(\beta - \alpha)/2}{\tan\beta} + \sin\left(\frac{(\beta - \alpha)}{2}\right)\right]. \tag{58}$$

It can be assumed, especially at the beginning of the collapsing cone in Fig. 41, that $\alpha = \beta$, whereupon Eq. 51 becomes

$$\begin{aligned} V_j &= V_o(1 + \cos\alpha)\sin\alpha \\ V_s &= V_o(1 - \cos\alpha)\sin\alpha. \end{aligned} \tag{59}$$

Equation 59 characterizes the velocity of the jet tip, while at the end of the cone collapse, $\alpha \to 0$, and V_j becomes very small (or a minimum velocity at the back of the jet, $V_{min.}$). In addition, the jet mass (ρ_j x ℓ_j x π $d_j^2/4$) (Figs. 40a and 41) becomes

$$m_j = m\ \sin^2(\alpha/2). \tag{60}$$

As an antitank or related military weapon, and for other applications such as well fracturing in oil and gas exploration and well drilling, the important shaped charge property is penetration, into either a target or a rock mass. For armor penetration, the formation of the jet prior to penetration and the integrity of the jet, that is, the particulation and contiguity of the jet particles, is a paramount consideration. Jet formation prior to penetration requires detonation and cone collapse corresponding to a standoff distance, S, from the target surface as illustrated in Fig. 42. Ideally, penetration can be approximated by

$$p \cong S\left[\left(V_{jtip}/V_{min}\right)^{1/\gamma} - 1\right], \tag{61}$$

where V_{jtip} is the tip velocity given by Eq. 59 for V and V_{min} is the minimum velocity to penetrate the target, and

$$\gamma = \left(\rho_t/\rho_j\right)^{1/2}. \tag{62}$$

Equation 61 applies to the unparticulated or contiguous jet. When the jet breaks up or particulates,

$$p \cong \left[\left(V_{jtip} - V_{min.}\right)t_b\right]/\gamma \tag{63}$$

where t_b is the breakup time as given by

$$t_b = d_j / V_{pl} \tag{64}$$

and

$$V_{pl} = \left(V_{jtip} - V_r\right)/n, \tag{65}$$

where V_r is the rear jet velocity and n is the number of jet particles shown in Fig. 40a. Equations 61 and 63 can be approximated by the hydrodynamic approximation where for the unparticulated (contiguous) jet,

$$p \cong \ell_j \left(\rho_j / \rho_t\right)^{1/2}, \tag{66}$$

which is the same form as Eq. 22, and ℓ_j is the jet length as shown in Fig. 1a, or for the particulated jet,

$$p \cong \ell'_j \left(2\rho'_j / \rho_t\right)^{1/2} \tag{67}$$

where ρ'_j is the estimated jet density including the open spaces between jet particles and ℓ'_j is the length of the particulated jet.

In retrospect, it can be observed that the equations describing shaped charge cone operation and jet penetration are both complex and generally approximate. The jet breakup is similar to fluid stream breakup and is dependent upon the DRX process illustrated schematically in Fig. 40b, which also involves dynamic ductility or superplastic deformation. It can be noticed that in the equations describing the jetting and penetration process, there is no indication of this material property. However, experimental measurements of penetration, as it relates to the explosive charge (or cone) diameter shown in Fig. 1a at CD, provide some indication of liner properties: apex (or cone) angle (α or 2α in Fig. 1a), density, grain size, and static ductility or elongation. These performance issues are illustrated in Fig. 43. It can be observed in Fig. 43 that with any variation in the liner apex angle, 2α (or α, the half angle), the liner (and corresponding jet density), the standoff distance (S = N(CD)), or the liner grain size, there is some penetration optimum (p). Consequently, there are a range of interactive design parameters for efficient or specific shaped charge performance requirements. An optimum apex angle (2α) occurs around 42°, which has been a US Army standard for 81 mm (CD) copper shaped charge liner cones, which as shown in Fig. 43 exhibits good penetration performance.

It can be observed from Eq. 55 that in the limiting case where $\alpha = \beta = 0$, $V_o = 2V_D$. Consequently, high explosives assure optimum velocities (>15 km/s) in the hypervelocity range, with the minimum or rear jet velocities only one-tenth of the jet-tip velocity.

As discussed in connection with liner flow in Fig. 41, the inner and outer liner regimes flow into the stagnation point (M in Fig. 41) by DRX. For a liner having an apex angle (2α) of 42°, the jet mass (m_j) according to Eq. 60 becomes $m_j \cong 0.2$ m. Correspondingly, the outer 4/5 of the liner mass flows into the slug production:

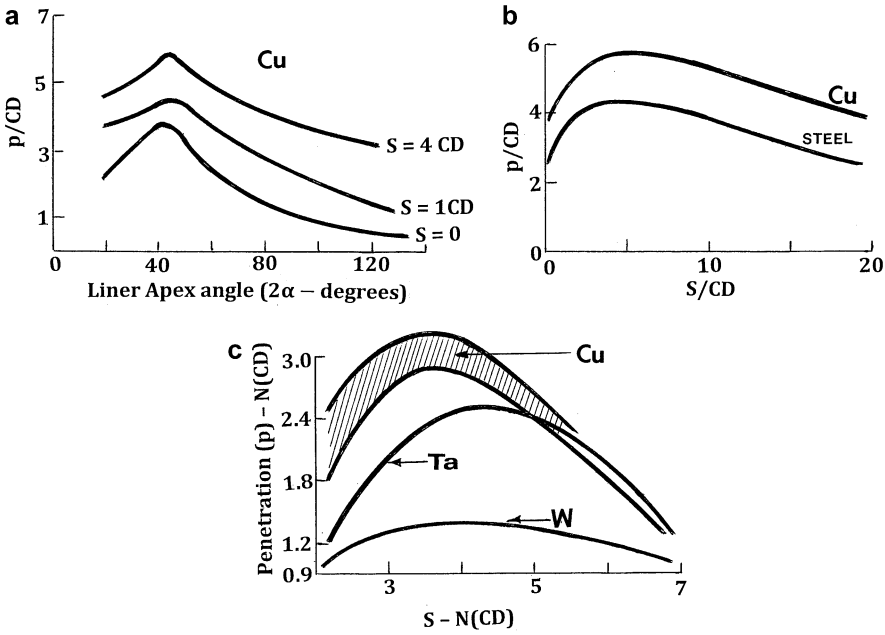


Fig. 43 Conical shaped charge performance plots. (**a**) Jet penetration for Cu liners with CD = 9.8 cm using octol HE (HMX/TNT mixture; $\rho_e = 1.8$ g/cm^3). The target is mild steel: $\rho_t = 7.8$ g/cm^3. Plots for different standoff (S) are shown. (**b**) Penetration to charge diameter ratio (p/CD) versus standoff (S) to charge diameter (CD) ratio (S/CD) for copper and steel jets impacting steel targets. Apex angle (2α) = 42°. (**c**) Standoff distance (S) = N (CD); N = multiple versus penetration depth (p) into steel armor targets; P = N (CD), for different liner materials. The shaded Cu region characterizes decreasing liner grain sizes (from ~105 to <10 μm). The penetration increases with decreasing liner grain size

$m_s \cong 0.8$ m. This produces a jet often having d_j (Fig. 41) of ~2 mm, or the diameter of a small straw, which can produce a penetration cavity having a diameter of >5 times this diameter in a steel target. Figure 44a illustrates a soft-recovered Ta jet fragment (from a Ta shaped charge liner), while Fig. 44b shows a section view of the equiaxed grain structure in a recovered Ta jet particle. Since the jet is significantly heated by adiabatic DRX, the residual jet temperature (>800 °C for Ta) allows for rapid grain growth of the DRX structure which permits solid-state flow of liner material into the jet as well as the superplastic elongation of the jet. For the slug (~80 % of the liner mass), the temperature rise is much smaller, and the flow-related deformation more prominent. This is illustrated in the sequence of microstructure images shown in Fig. 45. Figure 45a shows the initial Ta liner, equiaxed grain structure, while Fig. 45b shows the outer, heavily deformed slug microstructure to be composed of wavy, elongated cell-like features. Figure 45c shows the DRX, equiaxed grain structure in the center of the Ta slug in contrast to the starting Ta liner grain structure shown above in Fig. 45a. Correspondingly, Fig. 45d shows the elongated nano-cell structure characteristic of Fig. 45b.

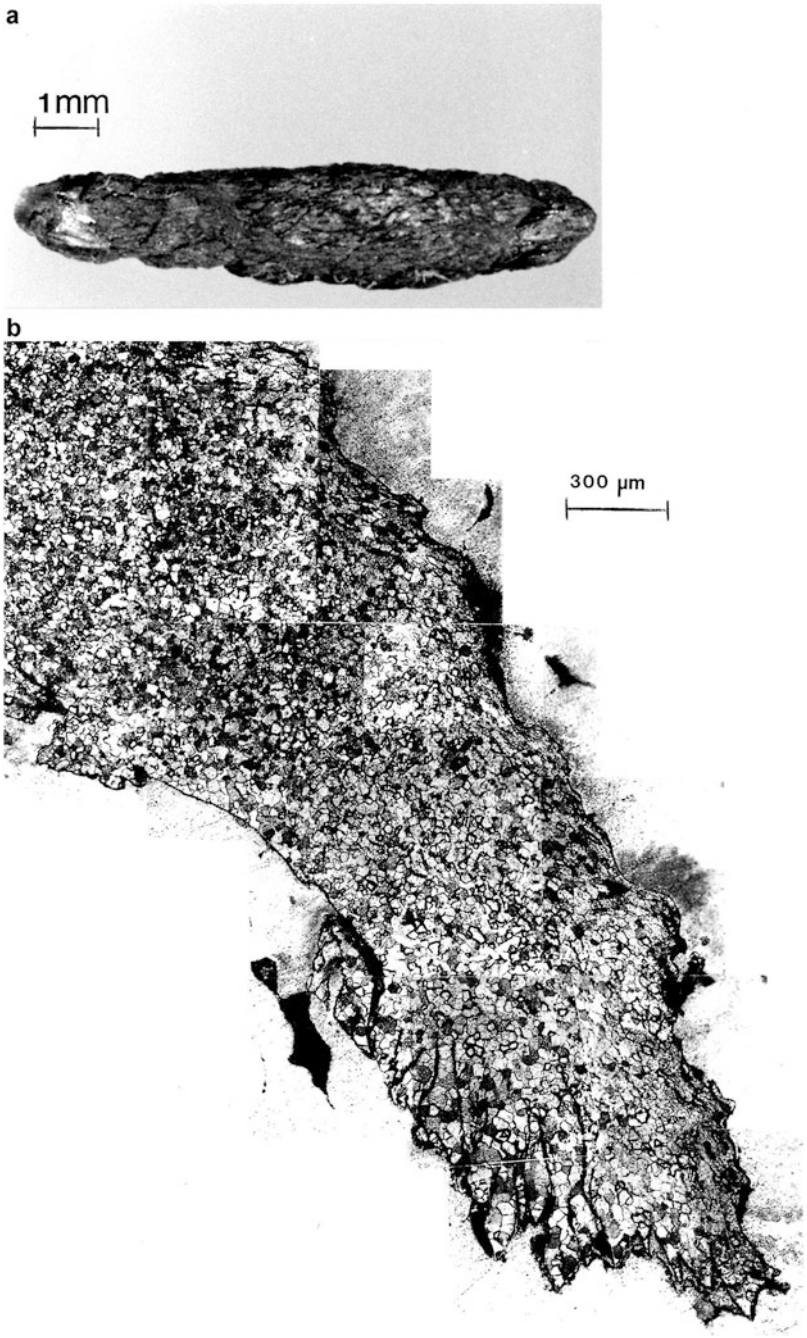


Fig. 44 Ta jet fragment for conical shaped charge. (**a**) Jet particle collected from particulated jet. (**b**) Equiaxed grain structure in jet fragment cross section

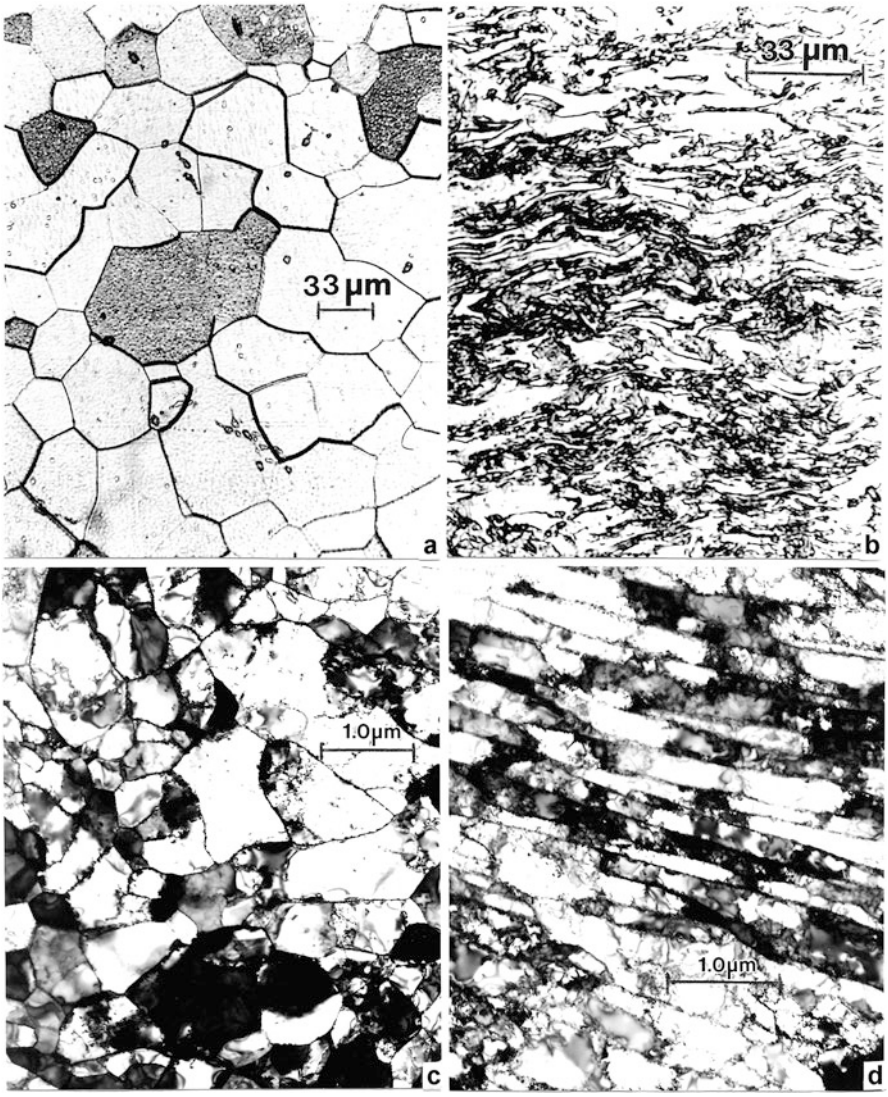


Fig. 45 Comparison of Ta liner grain structure (**a**) with outer slug deformation structure (**b**) and (**d**), with axial DRX grain structure in slug center. (**a**) Optical microscope image. (**b**) Optical microscope image. (**c**) and (**d**) are TEM images. (**b**) and (**d**) are corresponding microstructures

There are several variants of shaped charges. Figure 46a shows a tandem string of small, large apex angle shaped charges which can be accommodated in an oil or gas well pipe bore and lowered to a depth suitable for detonation to create a perforated rock regime conducive to inflow of oil or gas. Such shaped charge arrays account for hundreds of millions of shaped charges utilized in oil and gas production worldwide over the past decade alone. Figure 46b shows a linear shaped charge

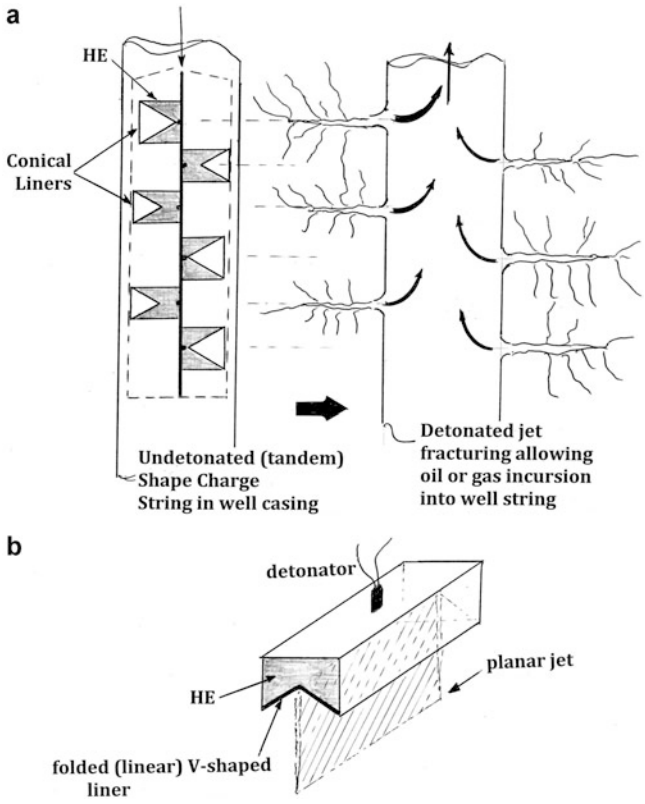


Fig. 46 Tandem array or shaped charge string for oil and gas flow enhancement by well-bore fracturing (**a**). (**b**) shows a linear shaped charge which produces a knifelike (planar) jet front upon detonation

variant which can produce a linear jet capable of cutting concrete, steel, or steel-reinforced concrete pillars or columns in demolition operations.

Figure 41b shows a self-forming projectile or explosively formed projectile (EFP) from a liner having a very large apex angle ($2\alpha \geqq 120°$) where the projectile formed is effectively a slug created from the complete liner mass. In contrast to the conical shaped charge in Fig. 41a, the EFP in Fig. 41c achieves velocities at impact of only 1.5–3 km/s. These velocities are approximately represented by applying Eq. 55, where ratios of M/C vary from $\sim$1.5 to 2.0. EFP formation must also involve a standoff from the intended target and usually involves distances at least a factor of 100 CD (charge diameter illustrated in Fig. 41c). Slower-moving EFPs against thinner targets will often produce a plug as illustrated in Figs. 36 and 37. Table 2 illustrates some experimental EFP characteristics for an Fe liner and corresponding standoff of 15 m from a steel target. Penetration of EFPs is often described by a hydrodynamic penetration equation having the form of Eq. 22:

Table 2 EFP performance (penetration) parameters for an iron liner referenced to Fig. 41c

CD (cm)	Liner mass M (kg)	Explosive mass C (kg)[a]	M/C	Penetration into steel[bρ (cm)]
7.6	0.52	0.27	1.9	2.5
15.2	4.3	2.5	1.7	7.0
20.3	8.7	5.1	1.7	10.0

[a]C-4 explosive (Table 1)
[b]Constant standoff, S = 15 m

$$\rho \cong \lambda \ell_s \left(\frac{\rho_s}{\rho_t}\right)^{1/2} \tag{68}$$

where λ is an efficiency factor which can vary from ~0.5 to 1.5, ℓ_s is the EFP (slug) length (Fig. 41c), ρ_s is the slug density, and ρ_t is the target density. Like conical shaped charges, EFP liners can be a variety of metals, Fe, Cu, Ta, etc., although Pb and Al liners have been used and are used in conical shaped charge arrays such as those illustrated in Fig. 46a. Because of its strength and ductility, Cu (as illustrated in Fig. 43) is a very efficient liner material for conical shaped charges. Depleted uranium has also been used in military HEAT applications, but is generally banned in modern warfare.

Material Issues in Railgun Performance

In the preceding sections of this chapter, ballistic and hypervelocity projectiles of various types have been discussed, especially in the context of military penetrators. These have involved various gunfirings using chemical propellants or various other explosive regimes and configurations. Rather conventional military guns for heavy penetrator launchings involve 120 mm bore guns which can fire multi-kilogram mass (~10 kg) projectiles up to ~2 km/s and distances up to 24 km. Guns of this type can be fired around 300 times before the barrel is damaged by erosion and corrosion. Larger naval battleship guns of the Iowa class (or 16 in. guns [406 mm bore diameter and 20 m long]) can fire 1,220 kg shells up to 39 km (24 miles) at a muzzle velocity of 820 m/s, using a powder charge of 300 kg. These systems were also limited in firings by thermal management issues, wear, erosion, and fatigue of the barrel. Consequently, there has been considerable work done in the development of coatings and claddings to prolong gun life especially for large guns.

During WWII, the Germans developed a series of even larger guns (variously referred to as Big Bertha, a gun built in WWI which could fire shells ~100 km). Figure 47 illustrates one such gun constructed circa 1941 with a 700 mm bore which could fire 5.5 m long armor – piercing shells weighing 7 metric tons more than 30 km. High-explosive shells could form a crater of ~30 m across and 10 m deep upon impact. As illustrated in Fig. 47, the gun required twin (parallel) railway tracks to transport it. It weighed around 1,400 metric tons and stood four stories tall. The barrel wore out after only 100 firings, which required 500 men to aim, fire, and

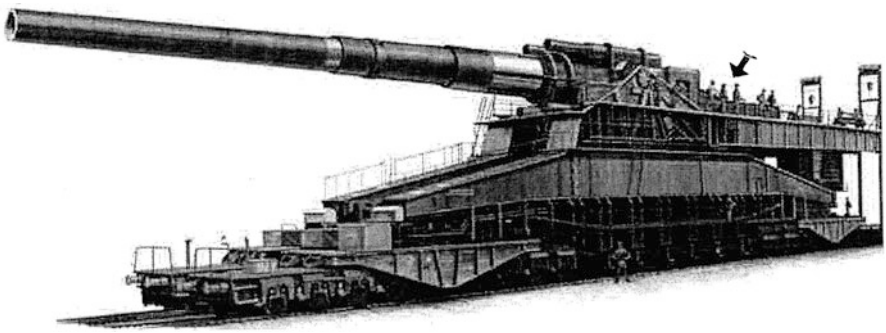


Fig. 47 WWII German-built supergun (Photograph courtesy of Railway Gun WEB Museum)

reload. Following WWII, a Canadian gun designer named Gerald Bull, working for the US Government on a supergun made by welding two 16 in. battleship guns together, end-to-end, was able to launch so-called "arrow" shells or finned shells having the appearance of small rockets to heights exceeding 200 km, lower Earth orbit. This prompted the prospects for launching satellites into Earth orbit using large guns.

Around the end of WWI, a French inventor, Louis Octave Fauchon-Villeplee developed an electric cannon as an early form of the modern railgun: "an electric apparatus for propelling projectiles" as his patent was titled. At the end of WWII, the Germans built the first working railgun. Since then, the development of single-pulsed homopolar generator-inductor railgun systems has been ongoing.

The simplest railgun design configuration consists of two parallel, conducting metal rails connected to an electrical pulse-power source with a conductive projectile inserted between the rails to complete the circuit as illustrated in Fig. 48a. This operates as an electromagnetic gun or projectile launching system that accelerates the conductive projectile along the rails using the same principle as the homopolar motor, with the contacting projectile acting as an armature on the stator (or static) rails. The principle of projectile launch involves a very high current (pulse current, I) passed through the rails and the projectile armature (Fig. 48a). The projectile current interacts with the strong magnetic fields generated by the rails to produce a correspondingly strong, pulsed force, the Lorentz force (F), where from Fig. 48a

$$F = \left(\frac{\mu_{o} I^{2}}{\pi}\right) \ln\left(\frac{d - r}{r}\right), \tag{69}$$

and M_o is the magnetic permeability of free space. Ignoring friction and air drag on the projectile, the projectile acceleration is given by

$$a = dv/dt = L' I^{2}/2m, \tag{70}$$

where L' is the inductance at the position of the projectile along the rails. Both Eqs. 69 and 70 illustrate that in order to accelerate a large projectile mass at high

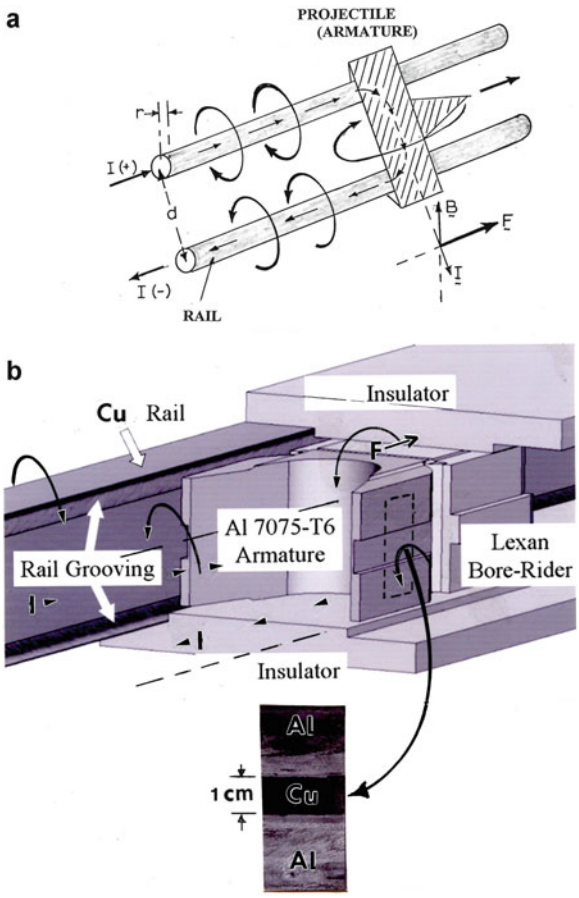


Fig. 48 Schematic (**a**) and section view (**b**) for railgun principles and design features. In (**a**) the armature (projectile) field interaction with the rail magnetic flux creates an accelerating force (F) on the projectile. (**b**) shows a railgun section after Watt et al. (2007). An Al–Cu–Al erosion surface section is shown removed from the experimental structure. (**a**) is after Machado et al. (2011)

velocity requires extremely high pulsed launch currents of at least a few million amperes. Muzzle velocities between 3 and 3.6 km/s for projectiles having a mass >2 kg would require muzzle energies near 50 MJ, with a launch current of ~10 MA for a few microseconds. This would require a large capacitor bank and DC generator system, or power supply, such as a nuclear reactor aboard a naval vessel.

Major contributions to railgun performance limitations are similar to those in large guns as illustrated in Fig. 47, but in addition, excessive rail resistive heating at large current pulses contributes to frictional heating between the rail and projectile during launch. The rails also experience a sideways force due to the magnetic field and their strength can be compromised by the elevated temperature. Rail erosion accompanies these features along with projectile ablation during each launch. The rails experience mechanical deformation below the eroded layer which also involves DRX which as in other ballistic or hypervelocity weapons systems facilitates solid-state flow. Figures 49 and 50 illustrate examples of DRX associated with Cu conducting rail flow and erosion as well as Al 7,075 alloy armature erosion

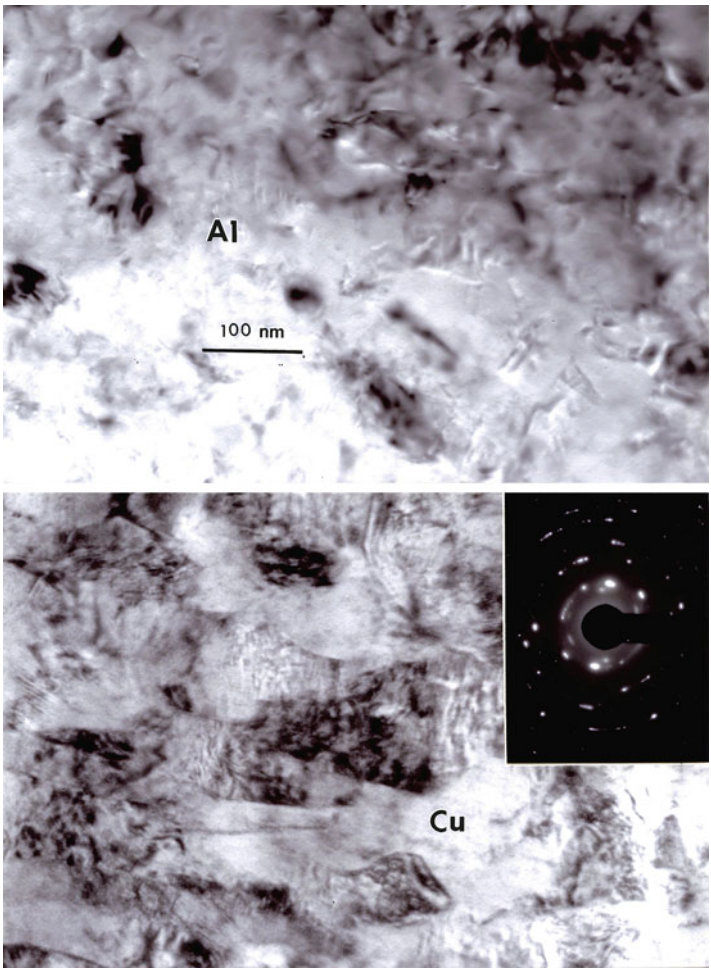


Fig. 49 TEM images showing DRX grain structure in Al and Cu erosion surface regions extracted from section shown in Fig. 48b (From Machado et al. 2011)

for an experimental railgun configuration shown in a cutaway view in Fig. 48b. The DRX grains in the Cu rail surface section in Fig. 49a and the Al armature projectile in Fig. 49b also exhibit a high degree of deformation vested in a high dislocation density. This is also evident in the interfacial (Cu/Al) regime shown in Fig. 50 where high-resolution TEM shows a dislocation density $\sim 10^{11}$ cm^{-2}. However, these microstructural features are obliterated under more extreme operating conditions which can lead to melting and associated plasma erosion (Watt et al. 2007). These phenomena are implicit in actual railgun test firing as illustrated in Fig. 51 which shows a 2008 US Navy test firing of a railgun at 10.6 MJ with a muzzle velocity for the projectile of 2.52 km/s. Later tests in 2010 and 2012 involved high-energy capacitor banks delivering 33 MJ pulsed energy.

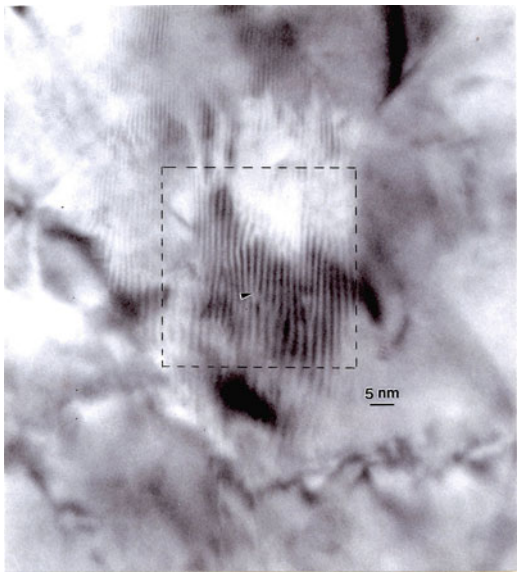


Fig. 50 High-resolution TEM image in DRX grains in Al–Cu interfacial region for extracted erosion section in Fig. 48. *Arrow* shows a prominent dislocation within a dotted area measuring 25 nm on a side (625 nm^2) (From Machado et al. 2011)

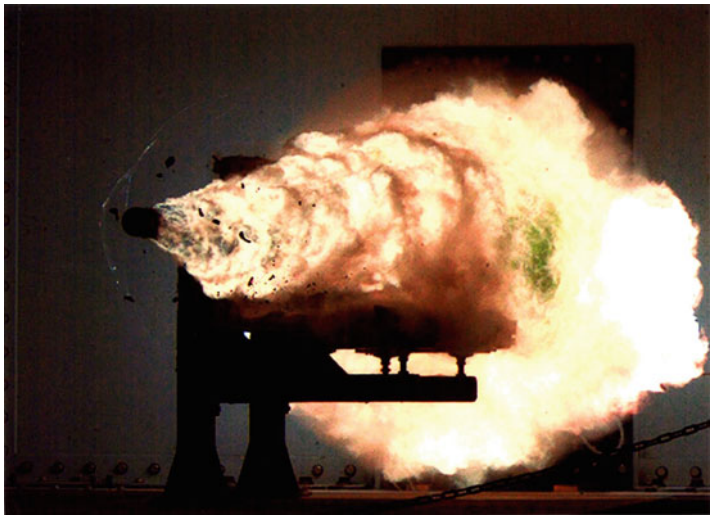


Fig. 51 Test firing of a railgun at the US Naval Surface Warfare Center in January, 2008, showing plasma plume behind the accelerating projectile (Courtesy U.S. Department of Defense, U.S. Naval Surface Warfare Center)

As noted in connection with large powder guns illustrated in Fig. 47, prospects for launching satellites into space have been considered for nearly a half century. Electrodynamic launch assistance has also been considered for railguns, but such launch applications would require rails of more than 1,500 m in length and estimated muzzle velocities >7 km/s at energies >30 GJ. Such conditions would

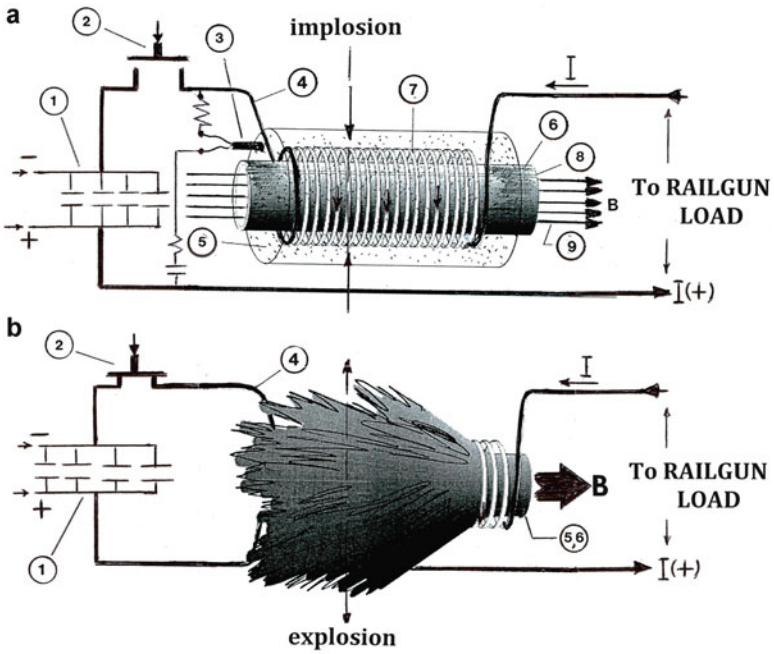


Fig. 52 Explosively driven magnetic flux compression generator concepts. (**a**) MK-1 implosive concept schematic. (**b**) MK-2 explosive concept schematic. (*1*) shows high-energy capacitor bank. (*2*) is activation (pulse) switch. Open in (**a**) and closed in (**b**). (*3*) Detonator with slight delay. (*4*) Current-carrying conductor to solenoid. (*5*) HE (high explosive). (*6*) Conducting (Cu) cylinder: stator in (**a**), armature in (**b**). (*7*) Solenoid flux coil. (*8*) Small slit in the Cu cylinder to promote flux propagation with explosive-initiated closure

not be tractable with conventional metals such as Cu for rail use, and some novel composite material systems would have to be developed. In addition, such enormous pulse-power requirements might only be provided by some form of *magnetic flux compression generator* or explosively driven magnetic flux generator, which could provide a one-shot destructive pulse launch of a heavy armature projectile. Such pulse-power systems also demonstrate extreme material deformation as illustrated schematically in Fig. 52 which shows two popular versions for *explosive flux compression generation*: implosion-initiated solenoid armature (Fig. 52a) and explosion-initiated conducting cylinder armature (Fig. 52b). In these systems, extremely high current–magnetic flux is pulsed into a large Cu solenoid coil with a small detonation delay following the power pulse from a bank of high-energy storage capacitors. The entire device could ideally be assembled in a chain of successive stages requiring increasing conductor sizes to accommodate the enormous, amplified energy pulses. In Fig. 52, the explosive detonation produces a convergent cylindrical shock wave which compresses or squeezes the magnetic field (flux) into an intense flux concentration which can attain field strengths $>$200 T and load currents $>$100 MA.

Flux compressor systems as depicted in Fig. 52 also have other pulse-power applications such as high-power laser pulsing, microwave pulsing, and related electromagnetic pulse (EMP) production. Since these systems are explosively destroyed, their actual use in such applications has some practical limitations.

References

Charters AC (1960) High speed impact. Sci Am 203(4):128–140
Christiansen EL (2009) Handbook for designing MMOD protection, NASA/TM-2009021478. NASA Johnson Space Center, Houston
Christiansen EL, Kerr JH (2001) Ballistic limit equations for spacecraft shielding. Int J Impact Eng 26:93–104
Dodd B, Bai YI (1989) Width of adiabatic shear bands formed under combined stresses. Mater Sci Technol 5:557–559
Esquivel EV, Murr LE (2006) Deformation effects in shocked metals and alloys. Mater Sci Technol 22(4):438–452
Grace FI (1995) Long-rod penetration into targets of finite thickness at normal impact. Int J Impact Eng 16:419–433
Grace FI (2001) Energy deposition during rod penetration in multiple-layered targets of steel and titanium. In: Staudhammer KP, Murr LE, Meyers MA (eds) Fundamental issues and applications of shock-wave and high-strain-rate phenomena. Elsevier, New York, pp 353–360
Grady DE, Kipp ME (1998) Impact failure and fragmentation properties of metals. Sandia Report, SAND 98-0387, March, 1998, Sandia National Laboratories, Albuquerque, NM
Hernandez VS, Murr LE, Anchondo IA (2006) Experimental observations and computer simulations for metallic projectile fragmentation and impact crater development in thick metal targets. Int J Impact Eng 32:1981–1999
Johnson GR, Cook WH (1983) A constitutive model and data for metals subjected to large strains, high strain rates, and temperature. In: Proceedings 7th international symposium on ballistics, The Hague, 1983
Machado BI, Murr LE, Martinez E, Gaytan SM, Satapathy S (2011) Materials characterization of railgun erosion phenomena. Mater Sci Eng A 528:7552–7559
Meyers MA (1994) Dynamic behavior of materials. Wiley, New York
Meyers MA, Murr LE (1981) Appendix C to F. In: Meyers MA, Murr LE (eds) Shock waves and high-strain-rate phenomena in metals. Plenum, New York, pp 1045–1078
Murr LE (1975) Interfacial phenomena in metals and alloys, Addison-Wesley Publishing Co, Inc, Reading, MA
Murr LE (1981) Residual microstructure-mechanical property relationships in shock-loaded metals and alloys. In: Meyers MA, Murr LE (eds) Shock waves and high-strain-rate phenomena in metals. Plenum, New York, pp 607–672
Murr LE (1988) Shock hardening and strengthening. In: Murr LE (ed) Shock waves for industrial applications. Noyes Publications, Park Ridge, pp 60–102
Murr LE (2012) Correlating impact related residual microstructures through 2D computer simulations and microindentation hardness mapping: review. Mater Sci Technol 28(9–10): 1108–1126
Murr LE, Hare AW (1988) Explosive forming and material working applications. In: Murr LE (ed) Shock waves for industrial applications. Noyes Publications, Park Ridge, pp 103–136
Murr LE, Pizana C (2007) Dynamic recrystallization: the dynamic deformation regime. Metall Mater Trans A 38A:2611–2621
Murr LE, Ramirez AC, Gaytan SM, Lopez MI, Martinez EY, Hernandez DH, Martinez E (2009) Microstructure evolution associated with adiabatic shear bands and shear-band failure in ballistic plug formation in Ti-6Al-4V targets. Mater Sci Eng A 516:205–216

Pappu S, Sen S, Murr LE, Kapoor D, Magness LS (2001) Deformation twins in oriented, columnar-grained tungsten rod ballistic penetrators. Mater Sci Eng A298(1–2):144–157

Quinones SA, Rivas JM, Garcia EP, Murr LE, Hörz F, Bernhard RP (1995) Microstructure evolution associated with hypervelocity impact craters in OFHC copper. In: Murr LE, Staudhammer KP, Meyers MA (eds) Metallurgical and materials applications of shock-wave and high-strain-rate phenomena. Elsevier, New York, pp 293–302

Reid SR, Wen HM (2000) Perforation of FRP laminates and sandwich panels subjected to missile impact. In: Reid SR, Zhou G (eds) Impact behavior of fibre-reinforced composite materials and structures. Woodhead Publishers, Cambridge

Valerio-Flores OL, Murr LE, Hernandez VS, Quinones SA (2004) Observations and simulations of the low velocity-to-hypervelocity impact crater transition for a range of penetrator densities into thick aluminum targets. J Mater Sci 39:6271–6289

Watt T, Stefani F, Crawford M, Mark H, Parker J (2007) Investigation of damage to solid-armature railguns at startup. IEEE Trans Magn 43(1):214–218

Watts A, Atkinson D, Ricco S (1993) Dimensional scaling for impact cratering and perforation. NASA contractor report NCR-188259, 16 Mar 1993. POD Associates, Albuquerque

Whipple FC (1947) Meteorites and space travel. Astronomical J 1161:131–135

Explosive Welding, Forming, and Powder Consolidation

Contents

Abstract

The use of explosives to manipulate and process materials epitomizes extreme deformation processing and provides additional evidence for the role played by dynamic recrystallization or related solid-state flow induced by strong shock waves. The explosive bonding or cladding of large areas of planar metal or alloy laminates by variations of interlocking weld waves or shallow DRX microstructures often resembles rim structures characteristic of ballistic impact craters and penetration channels. Novel forming by explosive charges in water creating hydroforming environments is presented. The explosive consolidation of powder materials can provide novel cylindrical precursors for more conventional materials processing or new materials fabrication not tenable by conventional materials processing routes.

Introduction

Roughly 10 years after John Rinehart first outlined prospects for using explosively generated shock-wave phenomena in the manufacturing of metal parts, he and John Pearson chronicled a host of novel and practical applications of shock waves in forming and joining a variety of metals and alloys (Rinehart and Pearson 1963). Similar *explosive metal working* was simultaneously initiated and developed during

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_50

World War II in the USSR (Novosibirsk) as chronicled by Deribas (1981). In addition to welding large, multiple metal and alloy plates (laminates) over their area of contact, and the free and die forming of metal plates using explosives, explosive sheets were detonated on large metal castings and forgings to notably harden the surfaces. Similar explosive sheet configurations were also utilized by DeCarli and Jamieson (1961) to synthesize diamond powder from graphite. Around this time period, concerted efforts were launched in the USSR to develop effective powder compaction and consolidation technologies using explosively generated shock waves in a variety of configurations, at pressures up to 100 GPa (Roman and Gorobtsov 1988). Explosive shock compaction has also been used in novel shock-induced chemical reactions of powdered materials and the production of immiscible, high-temperature alloy billets for plate and extrusion-product fabrication.

Explosive Welding/Cladding/Bonding

Explosive welding, also referred to as explosive bonding and cladding, especially over the impacted surface for two or more parallel plates of metals and/or alloys (planar geometries), has been commercialized for at least a half century or more. The concept, involving planar geometries or explosive welding of parallel metal plates, is illustrated schematically in Fig. 1. This arrangement involves standing a flyer plate of thickness h_f and density ρ_f and a distance S (the standoff distance) from the target or base plate having a thickness h_b and density ρ_b. An explosive of height h_e and density ρ_e is placed over the flyer or cladding plate and either crudely detonated with a point detonation or used with a line-wave generator to create a uniform detonation front at the left of the parallel-plate configuration as shown in Fig. 1a, b. In the constant standoff configuration in Fig. 1a, b, detonation causes the plates to collide and bond by creating a collision geometry shown in Fig. 1c. In this geometry configuration, the collision point velocity, V_C, is equal to the explosive detonation velocity, V_D : $V_C = V_D$, and this implies that V_C is independent of the standoff distance, S. The turning angle of the flyer plate at the collision point at impact is equal to the collision angle or dynamic bend angle, β in Fig. 1c. The sine of this angle is roughly linearly related to the flyer-plate velocity, V_p:

$$V_C \cong V_D = \frac{V_p}{2\sin(\beta/2)} \cong \frac{V_p}{\sin\beta} = \text{constant}. \tag{1}$$

The maximum standoff can be approximated to be

$$S \cong \frac{3h_e}{4}\left(\frac{C}{M}\right) \tag{2}$$

where h_e is the explosive height shown in Fig. 1, and

$$C/M = h_e\rho_e/h_f\rho_f, \tag{3}$$

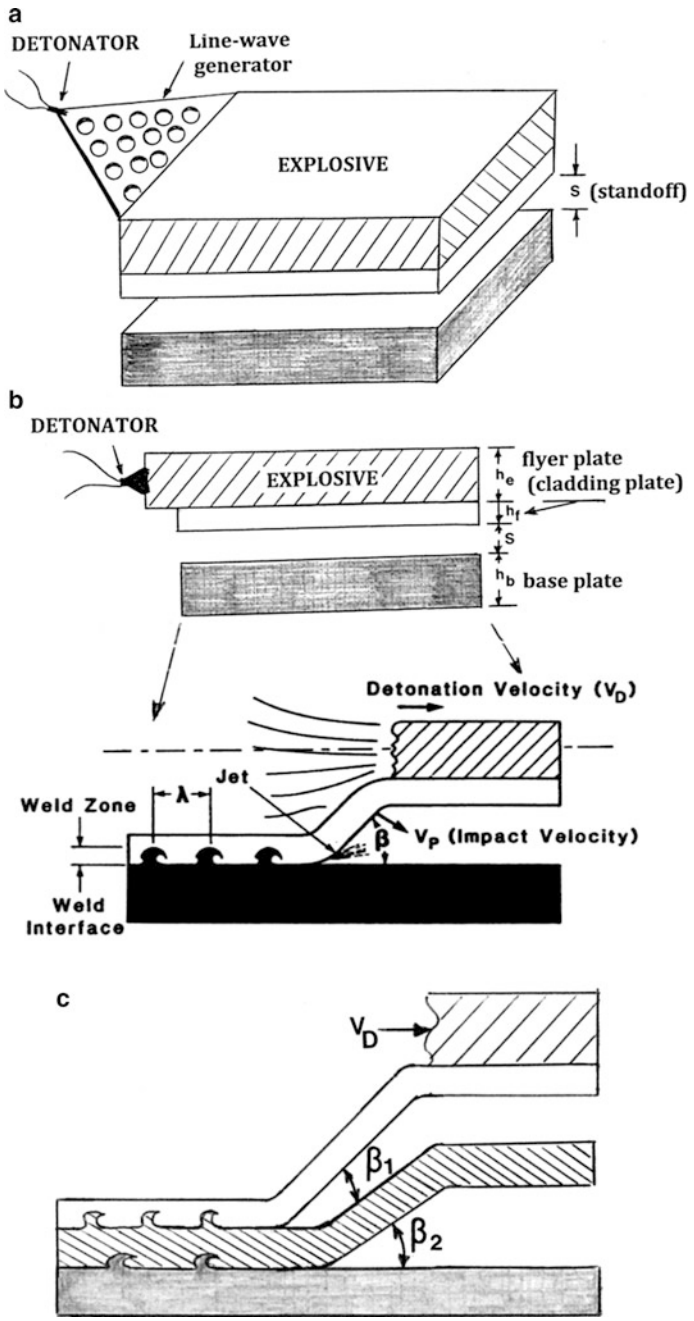


Fig. 1 Basic welding configurations. (**a**) Two-plate welding using plane-wave (line-wave) explosive generator. (**b**) Two-plate welding using point detonation of explosive charge. This mode of

where ρ_f is the flyer-plate density (Fig. 1). As a rule of thumb, welding of the plates usually requires $C/M \geq 2$. However, the standoff simply needs to allow the flyer time to achieve some minimum plate velocity for welding, which can be approximated by

$$V_p(\min) \cong (UTS/C_0)^{1/2}, \tag{4}$$

where UTS is the ultimate tensile strength and C_0 the bulk sound velocity for the flyer plate. This condition often allows a much smaller standoff than given in Eq. 2 to achieve a good weld (or bond), often $S' = 0.02$ S, where S' is the actual standoff. This standoff is usually achieved using small shims at the corners of the flyer and base plate, with a few edge shims for very large, thin, flyer plates (>2 m on a side; $h_f < 1$ cm for steel plates). The minimum velocity in Eq. 4 indicates that a minimum pressure is required at the collision point to create deformation producing dynamic recrystallization (DRX) and jetting of the metal surfaces to create a solid-state, interfacial flow and mixing to create a bond or surface weld.

This turbulent flow regime can create vortex or wave structures as shown in Fig. 1c having an amplitude characteristic of the weld zone and a displacement or wavelength, λ, where

$$\lambda \cong K\, h_f \sin^2(\beta/2). \tag{5}$$

The weld interface height or amplitude in Fig. 1c is also given by a similar equation, but the constant, K, is different (smaller).

It can be observed that a maximum flyer-plate velocity might be defined above where the pressure would create melting and poor or no bonding. An estimate for this velocity is given by (Chadwick and Jackson 1983):

$$V_p(\max) \cong \frac{K'(T_m C)^{1/2}}{V_D}\left(\frac{K_T C H}{\rho_f h_f}\right)^{1/4}, \tag{6}$$

where K' is a constant, C is the bulk sound velocity in the flyer plate, T_m is the melting temperature of the flyer plate, K_T is the thermal conductivity, and H is the specific heat. It can be observed in Eq. 6 that this velocity declines as the flyer density and thickness increase and with an increase in the explosive (detonation) velocity.

Fig. 1 (continued) detonation can create edge or end effects which are often cut from the final, welded product. Detonated plate impact forming weld interface characterized by weld-wave structure. Wave spacing is noted by λ (wavelength). (**c**) Triplex (multilayer) laminate explosion weld schematic

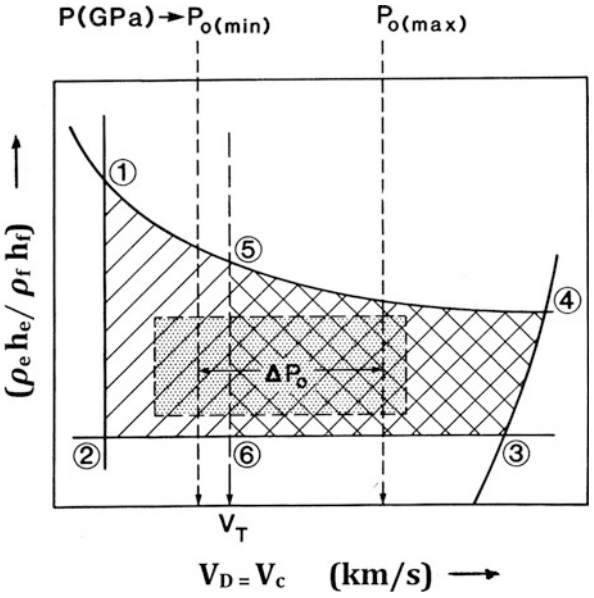


Fig. 2 Schematic illustration of explosion welding window: *1–4*. Transition velocity is denoted V_T; *5* and *6* optimum pressure regime is denoted ΔP_0; $\Delta P_0 \cong \rho_e \Delta V_D^2/4$. Weld waves decline below V_T and from *5* to *6*. Optimum window is shown in *rectangular shading zone*

For many explosive welding applications, low-cost ammonium nitrate/fuel oil mixtures are popular, even though V_D is much lower than composition C explosives (~1.5 vs. ~7 km/s). For these low-detonation-velocity, granular explosives, the flyer-plate velocity is reliably approximated by

$$V_p \cong 1.2\, V_D \left(\frac{(1 + 1.2R)^{1/2}}{(1 + 1.2R)^{1/2}} \right), \tag{7}$$

where $R = C/M = (\rho_e h_e/\rho_f h_f)$. Consequently, it can be observed on examining Eqs. 4 and 6, as well as Eq. 5, that a welding window plot or diagram can be constructed as illustrated in Fig. 2. The transitional velocity, V_T, in Fig. 2 denotes a minimum velocity above which a wavy weld interface occurs:

$$V_T = \left[\frac{2R_c(H_f + H_b)}{(\rho_f + \rho_b)} \right]^{1/2}, \tag{8}$$

where H_f and H_b are the Vickers microindentation hardness values for the flyer plate and impact plate, respectively, and R_c is an equivalent Reynolds number (~11).

Figure 3 illustrates a wavy interface structure in an Al/Cu weld using an ammonium nitrate/fuel oil (ANFO) composition with $V_D \cong 1.7$ km/s. The relative position of this weld configuration within the weld window is shown shaded in Fig. 2. The arrow in Fig. 3 shows a microdendritic weld-wave structure indicative of melting in the wave vortex, while the wave and vortex microstructure composing

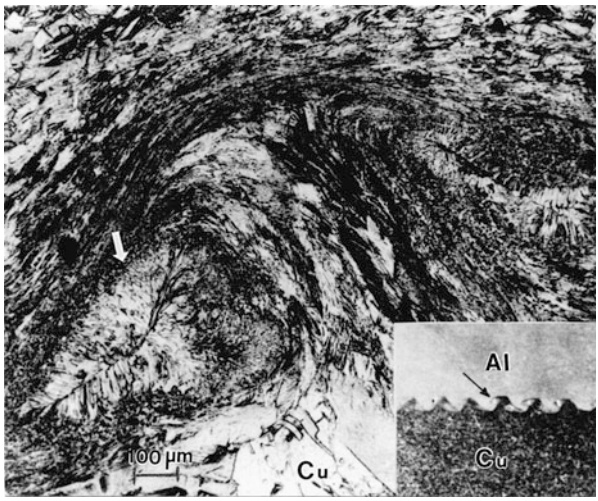


Fig. 3 Weld-wave and weld interface structure in Al/Cu explosion weld. *Insert* shows the explosion weld interface and weld-wave spacing. Note wave is reversed in *insert*

the Al/Cu weld region is characteristic of DRX solid-state flow. The relatively undeformed Cu grain structure is shown in the lower left of the weld wave indicated by the arrow in Fig. 3. It is especially noteworthy to compare the DRX weld interface structure in Fig. 3 with the DRX flow features for the ballistic penetration channel rim structure in a Cu target illustrated in ▶ Fig. 30 in chapter "Ballistic and Hypervelocity Impact and Penetration".

While large, multiple or multilaminate explosive welding systems can be fabricated by the continual stacking of parallel plates separated by a standoff or space as shown schematically in Fig. 1, planar configurations are often explosively welded using a vertical arrangement illustrated schematically in Fig. 4. In this arrangement, two welded configurations are produced simultaneously using a cardboard spacer to interrupt the converging shock waves. In a sense, the configuration of Fig. 1 is simply turned 90°. However, this can be viewed more of a free body arrangement of welding in free flight. Nonetheless, there is some aspect of free flight even attached to the concept shown in Fig. 1. Salam and Al-Hassani (1981) have developed some fundamental descriptions for free flight explosion welding which are illustrated schematically in Fig. 5a, b. Additionally, the welding of two plates shown in Fig. 5b can be extended to multilayer welding as illustrated in Fig. 5c.

Considering the application of the principle of momentum conservation in Fig. 5a, b, and in a direction normal to the common plane of the welded plates, Salam and Al-Hassani (1981) arrived at the following equation relating the velocity of the two welded plates (V_{p2}) to the flyer or first plate velocity (V_{p1}):

$$V_{p2} = V_{p1}\left(\frac{h_1\rho_1}{h_1 + h_1\rho_2}\right)\left(\frac{\cos\left(\beta/2 - \beta'\right)}{\cos\left(\beta'/2\right)}\right). \tag{9}$$

The corresponding kinetic energy loss of the system due to impact and welding is also given by

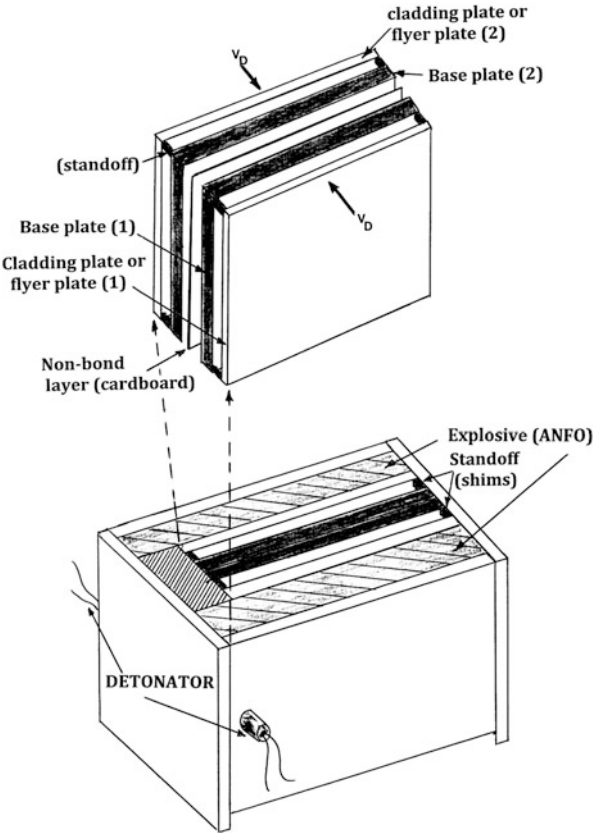


Fig. 4 Explosive fabrication arrays using two matched cladding systems

$$\Delta KE = \left(\frac{V_{p1} 2h_1\rho_1}{2}\right)\left[1 - \frac{h_1\rho_1}{h_1\rho_1 + h_2\rho_2}\left(\frac{\cos\left(\beta/2 - \beta'\right)}{\cos\left(\beta'/2\right)}\right)^2\right], \tag{10}$$

where $\beta' = \beta - \beta''$ in Fig. 5b. Considering the parallel welding configuration:

$$\left.\begin{aligned} V_{p1} &= 2V_D \sin\left(\beta/2\right) \\ V_{p2} &= 2V_D \sin\left(\beta'/2\right), \end{aligned}\right\} \tag{11}$$

Equation 10 can be written alternatively as

$$\Delta KE = 2\, V_D\left[h_1\rho_1\left(\sin^2(\beta/2) - \sin^2(\beta'/2)\right) - h_2\rho_2 \sin^2(\beta'/2)\right]. \tag{12}$$

Since this kinetic energy loss reflects the collisional plate impact and welding, the wave amplitude and the corresponding wavelength of the interfacial waves (weld interface) are proportional to it:

$$\lambda = K_1\, \Delta KE. \tag{13}$$

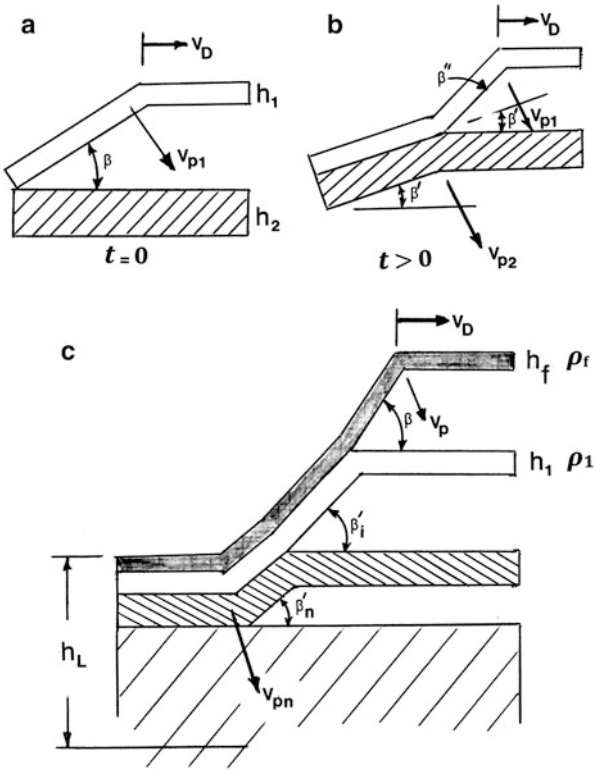


Fig. 5 Schematic views for explosion welding of plates in free flight. (**a**, **b**) show two plate configurations. (**c**) Multilaminate free flight schematic. V_D is the detonation velocity of the explosive charge

It can be observed that if in Eq. 12 $h_2\rho_2 >> h_1\rho_1$, then the configuration is essentially as shown in Fig. 1c, and if $\beta' = 0$, then

$$\Delta KE = 2\, V_D h_1 \rho_1\, \sin^2(\beta/2) \tag{14}$$

and Eq. 13 becomes

$$\lambda = K_1 2\, V_D h_1 \rho_1\, \sin^2(\beta/2), \tag{15}$$

which has the same form shown in Eq. 5, where $h_1 = h_f$ and $\rho_1 = \rho_f$.

For the multilaminate weld configuration shown in Fig. 5c, the system velocity is given by

$$V_{pn} = V_p \left(\frac{h_f \rho_f}{h_f \rho_{f+nh_L} \rho_L} \right) \left(\frac{\cos\left(\frac{\beta}{2} - \beta'_1\right) \prod_{i=1}^{n} \left[\cos\left(\frac{\beta'_{i-1}}{2} - \beta'_i \right) \right]}{\prod_{i=1}^{n} \left[\cos \beta'_i / 2 \right]} \right), \tag{16}$$

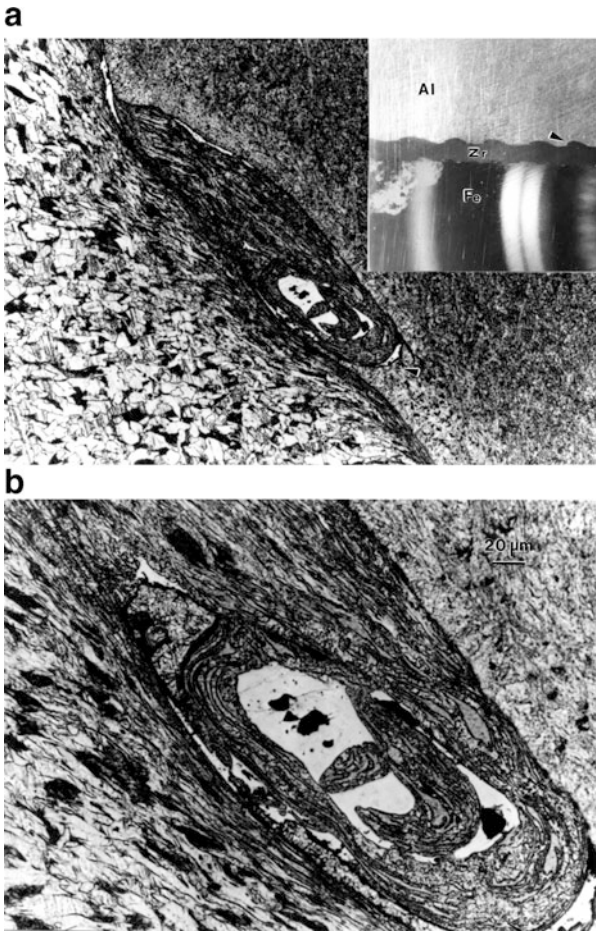


Fig. 6 Weld-wave and weld interface structure in aluminum/zirconium/steel triplex laminate shown by *insert* in (**a**). *Arrow* in *insert* shows Al/Zr weld wave. (**b**) Magnified image of (**a**)

where h_L and ρ_L are the complete welded layer thickness and density (average density), respectively, and n represents the total number of layers (including the flyer plate). In such complex configurations, there is considerable empiricism involved in the standoff or spacing between layers, and in solving Eq. 16, the velocity can be achieved by equating to Eq. 7 where $R = (h_e\rho_e/h_L\rho_L)$, and V_D can be adjusted using a suitable explosive.

Figure 6 shows a weld zone for a three-plate weld where a thin interlayer (Zr) is placed between aluminum and steel (Fe), $h_{Al}/h_{Zr}/h_{steel(Fe)}$: 0.13 cm/0.02 cm/2.54 cm. In contrast to Fig. 3, the weld wave in Fig. 6 does not show a microdendritic melt zone in the wave vortex, but a Zr-Al eutectic zone along with related reaction products occurs as a consequence of vortex-related elevated temperature. The weld interface is also characterized by DRX flow zones forming complex, fluidlike, and vortex-related microstructures similar to Fig. 3.

It might be noted that over the past half century, many thousands of explosive weld laminates have been fabricated, ranging in size (area) from < 1 to $> 20\ m^2$. Up to 20 different metal and alloy plates have been explosion welded into complex laminates. Laminates represented in Fig. 6 have been fabricated for many thousands of square meters of ship decking and related applications since the 1970s. Explosive cladding has produced many novel precursor laminates used to fabricate other geometries and related extensions of the technology which will be discussed in the following section.

It should also be noted that while this discussion has been mostly concerned with planar explosion welding, it is possible to weld many other geometries, including tubes, rods, and even plate geometries with initial, preselected angles. In addition, powder arrays have also been consolidated and incorporated in explosion welding (Murr et al. 1988; Murr 1991).

Explosive Forming Principles and Examples

Figure 7 compares conventional open die punch forming or stamping with closed and open die explosive hydroforming. The explosive die forming in Fig. 7d uses an explosive charge standing off axially from the forming blank a distance d_S. Figure 7e employs a shaped explosive charge to draw the forming blank into the dieless forming cavity. Both die-forming concepts illustrated in Fig. 7d, e rely upon an initial shock pressure generated by the explosive charge expressed generally by

$$P_o = K_o\left(\frac{W^{1/3}}{d_S}\right)^m, \tag{17}$$

where in water-forming systems such as those shown

$$P_o = \sqrt{6.9V_D}\left(\frac{W^{1/3}}{d_s}\right)^{1.13}. \tag{18}$$

The detonation velocity is shown as V_D in Eq. 18, while the explosive charge mass (weight) is given by W. The axial velocity at the center of the forming blank in Fig. 7d, e is given by

$$V_o = 2K_o'\left(\frac{W^{1/3}}{\rho_B t_B}\right)\left(\frac{W^{1/3}}{d_s}\right)^n, \tag{19}$$

where K_o' and n depend upon the nature of the explosive charge (V_D) and ρ_B and t_B are the forming blank density and thickness.

The forming time can be calculated from

$$\tau_F = 2\,h/V_o, \tag{20}$$

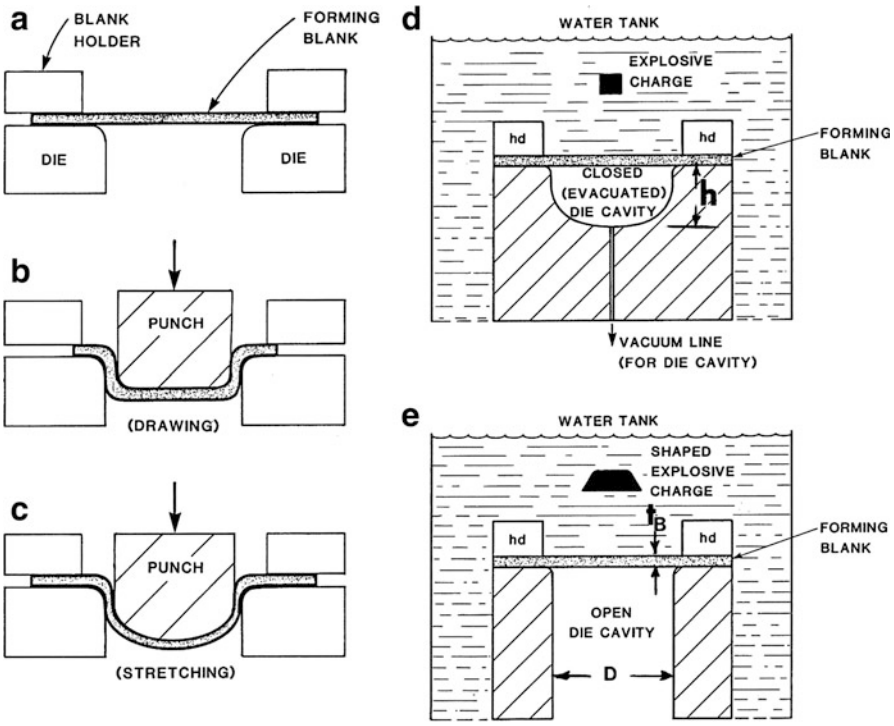


Fig. 7 Comparison of conventional forming principles (**a**–**c**) with explosive forming (in water) in a die cavity (**d**) and open (die) cavity (**e**). Holding clamps are denoted hd. Water acts as the explosive transfer medium (Adapted from Murr and Hare (1988))

where h is the distance from the initial position of the forming plate in Fig. 7d to the bottom of the die cavity. The corresponding strain rate along the explosively formed part in Fig. 7d will be expressed as

$$\dot{\varepsilon} = \varepsilon/\tau_f = \varepsilon V_o/2h \tag{21}$$

where ε is the strain. For the open die drawing of the forming blank in Fig. 7e, the draw distance can be approximated by

$$h_o = D\left(\rho_o/2\sigma_y' t_B\right)^{1/2}, \tag{22}$$

where D is the cavity diameter (Fig. 7e) and σ_y' is the dynamic yield stress, often approximated by the static yield stress (σ_y) : $\sigma_y' \cong \sigma_y$ up to $\dot{\varepsilon} \cong 10^3\ s^{-1}$.

Cylindrical configurations as shown in Fig. 8a can allow for tube-to-tube plate explosive welding or cylindrical (tube) forming (left to right, respectively). Tube expansion can be driven by an explosive cord, or the flyer tube (inner) can be filled with explosive. The tube will expand at the detonated end creating a forming angle,

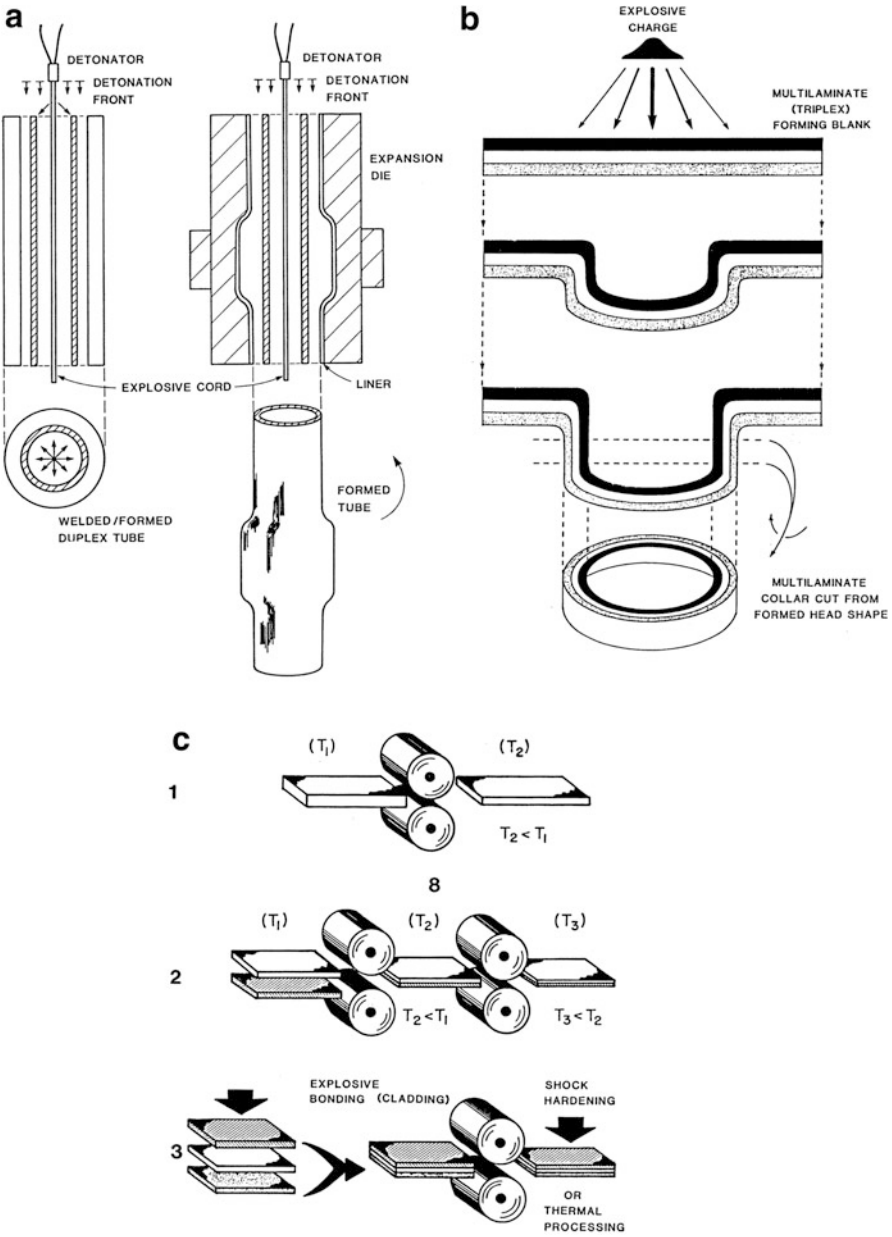


Fig. 8 Examples of explosive forming and metal working. (**a**) Cylindrical (tube) forming/welding. Forming die is shown at *right*. (**b**) Explosive forming or drawing of explosively clad plates to form novel, multilayer ring products. (**c**) Comparison of conventional rolling (*1*), roll cladding (*2*), and explosive cladding (welding) and rolling (*3*) (Adapted from Murr and Hare (1988))

θ_F, similar to the collision angle for planar explosion welding (Fig. 1c). The radial (tube) velocity can be approximated as

$$V_R = V_t = V_D \sin \theta_F. \tag{23}$$

Gurney (1947) has expressed the terminal tube velocity for an explosive-filled flyer tube as

$$V_t = \sqrt{2E}[R'/(1 + R'/2)]^{1/2}, \tag{24}$$

where E is the *Gurney energy* and

$$R' = \rho_e d_e / 2\rho_t h_t, \tag{25}$$

where ρ_e and d_e are the explosive density and diameter (tube inner diameter for filled tube) and ρ_t and h_t are the tube density and thickness, respectively. The collision or impact energy (E_I) is given by

$$E_I = M_t V_t^2 / 2\pi D_t', \tag{26}$$

where M_t is the tube mass and D_t' is the mean tube diameter at collision.

It can be observed in Fig. 8a that in adjusting the explosive (and detonation velocity, V_D), as well as the tube spacing (which adjusts for the forming angle, θ_F, in Eq. 23), the collision velocity can be adjusted to both form and weld (or bond the tubes) or simply form the flyer (inner) tube. In the configurations of Fig. 8a, the explosive cord does not create sufficient impact or collision energy for welding since the mass is small and is therefore an ideal configuration for cylindrical (tube) forming.

Figure 8b, c shows two additional forming processes involving explosives. In Fig 8b, either an explosively welded duplex or a multilaminate, parallel plate system is free-formed using a shaped explosive charge into a drawn cup geometry from which a novel, multilaminate collar can be cut as shown. This configuration could also be held in a forming arrangement as shown in Fig. 7b.

Figure 8c depicts conventional rolling or plate hot rolling (1), roll cladding followed by hot rolling (2), and explosive cladding, rolling, and final shock-wave hardening or other thermal processing (3).

Explosive Shock-Wave Consolidation of Powder Materials

Powder metallurgy forms a very broad processing route for producing metal and alloy products or product precursors. A prominent processing strategy has involved hot isostatic pressing or HIPing, where powder metals or alloys are compacted and consolidated by applying a high static pressure to heated powders to produce monolithic configurations. In the production of some consolidated products by HIPing, especially for high-temperature powders, the high

temperatures required to produce near theoretical (100 %) densities may be so high that the HIPed product becomes fully annealed and as a consequence may require extensive reprocessing or *thermomechanical treatment* to render the required performance properties.

Shock or explosive consolidation, on the other hand, can produce monolithic products from powder precursors which have very high residual hardness and strength as a consequence of the shock-wave hardening which accompanies compaction and consolidation. While a number of configurations for explosive consolidation are possible, the most efficient is a cylindrical configuration where the consolidated powder forms a monolithic cylinder or similar geometry. Figure 9 illustrates this configuration which creates an imploding cylindrical shock wave which consolidates the precursor powder into a homogeneous, dense (>95 %) monolith. In the configuration of Fig. 9, statically compacted metal or alloy powder (or other powdered material, including ceramics or metal-ceramic mixtures) in a closed metal container (which can be any convenient metal or alloy, such as steel tubing with welded ends) is placed in the center of a larger container (marked C in Fig. 9) which is then filled with explosive. Granular, high explosive such as TNT or mixtures and ammonium nitrate/fuel oil (ANFO) mixtures are also compacted to a desired density (ρ_e) and a wooden wave shaper placed over the enclosed powder sample to be consolidated. The explosive can be formed around the wave shaper as shown in Fig. 9 or simply filled to some height above it by extending the container (C). A detonator is placed at the top or apex of the explosive as shown. Upon detonation (Fig. 10) the detonation front travels downward creating a cylindrical shock pressure against the container wall which acts as a flyer tube against the powder. The dotted section to the right of the powder axis in the configuration of Fig. 9 can be viewed as an open-faced sandwich similar to the explosive welding of plates shown in Fig. 1a, b, where the flyer-plate velocity is given approximately by Eq. 7 or the original Gurney form in one dimension (Fig. 1):

$$V_f \cong \sqrt{2E}\left(\frac{3R^2}{R^2 + 5R + 4}\right)^{1/2} \tag{27}$$

where

$$R = C/M = \rho_e h_e / \rho_f h_f \tag{28}$$

in the configuration of Fig. 9. In the 2D perspective shown in Fig. 9, Eq. 27 may be adjusted somewhat:

$$V_f' \cong 0.75\, V_f, \tag{29}$$

or

$$V_f \cong \sqrt{2E}[R(1 + R/2)]^{1/2}, \tag{30}$$

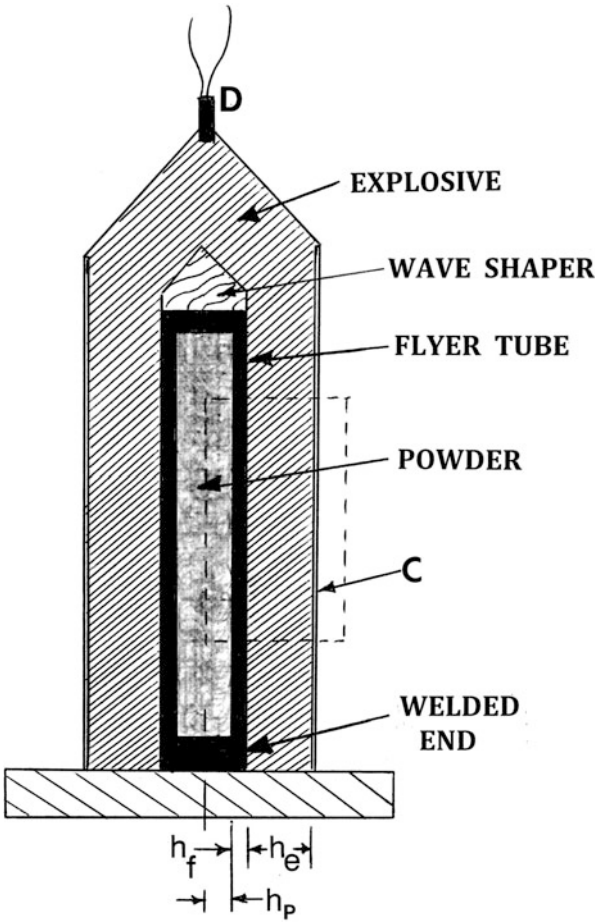


Fig. 9 Cylindrical explosive consolidation fixture for powder materials. *D* denotes detonation; *C* denotes the container which can be metal, plastic (*PVC*), cardboard, etc. The powder container is often a steel tube with welded steel ends. The bottom can be welded to the base which can also be metal and act as a shock-wave spall plate. The wave shaper can be wood or plastic. Explosive is granular and compacted to a desired density (ρ_e)

Fig. 10 Explosion event for large cylindrical consolidation of powder (Courtesy of Los Alamos National Laboratory)

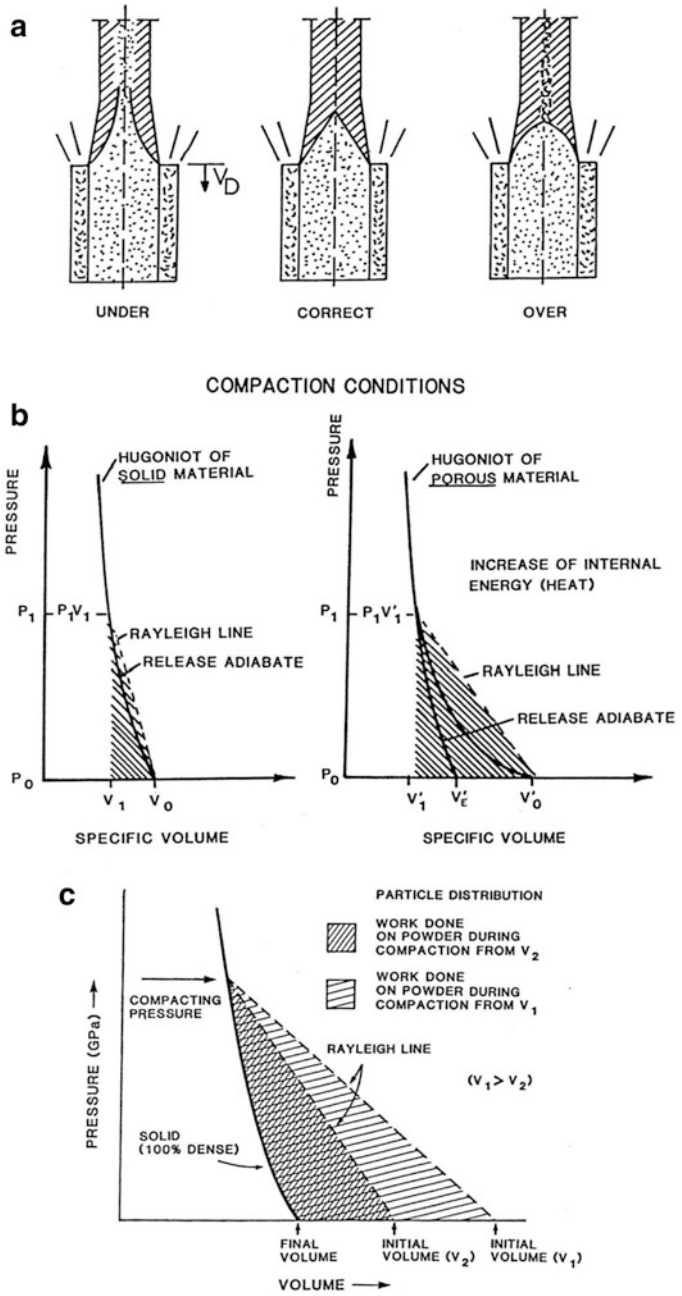


Fig. 11 Shock compaction/consolidation conditions (**a**) and Hugoniot pressure-volume relationships for metal powders (**b**). (**c**) shows details of the Hugoniot relationship for a powder having different initial packing densities (static compaction) (Adapted from Murr and Staudhammer (1988))

where E is the Gurney energy for the explosive (▶ Table 1 of chapter "Ballistic and Hypervelocity Impact and Penetration").

It is observed from Eqs. 27, 29, and 30 that the shock conditions or shock pressure conditions (shock pressure is proportional to V_f) are only approximate. In any case, trial and error runs are conducted over a range of explosive conditions (either variations in static compaction of the explosive to vary ρ_e or the variation of R in Eq. 28) in order to optimize the consolidated product: producing a homogeneous, maximum density cylinder. This process is summarized conceptually in Fig. 11 which shows poor compaction and consolidation for under-pressures (or flyer-tube velocities), the optimum (correct) pressure condition, and the over-pressure condition where the shock-wave convergence along the center of the powder tube creates a so-called Mach stem where overpressures can melt the center of the consolidated cylinder or create irregular consolidation. Similarly, under-pressures can cause poor or no consolidation along the cylinder axis or poor consolidation throughout the compacted powder cylinder. Figure 11b, c shows the corresponding pressure-volume (shock compressibility) relationships for a solid versus a porous (or variously compacted) material. Figure 11c in particular shows the Hugoniot (pressure-volume) relationship for a powder having different densities (or particle sizes or size distribution) as shown at V_1 and V_2. The line connecting the initial volume at zero pressure and the pressure corresponding to full density (100 %) is called the Raleigh line.

Figure 12 illustrates examples of precursor metal and alloy powders having very different pressure (flyer velocity) requirements as a consequence of very different melting points as well as mechanical properties, such as strength and hardness. The powders in Fig. 12 also illustrate variations in powder size and size distribution, which can contribute to both static compaction and explosive shock consolidation efficiency.

Figure 13 illustrates the explosive shock consolidation of a relatively narrow size distribution MAR-M200 (Ni-base superalloy) powder. The starting powder, a rapidly solidified (atomized) powder similar to those shown in Fig. 12, had an average size of ~50 μm. This narrow size distribution is reflected in the consolidated grain structure shown in Fig. 13d. In well-consolidated products, the shock wave will deform the particles, while a plastic wave can be considered to deform the contacting regions of the consolidating powder particles as shown in Fig. 13a, b. This creates a DRX zone which can vary with powder properties (strength, hardness, melting point etc.) and the explosive shock-wave parameters as shown in Fig. 13c. Figure 13e, f illustrates TEM microstructures corresponding to the consolidated MAR-M200 cylinder in Fig. 13d, as these relate to heavy deformation microstructures within the particle interiors (Fig. 13e) or regions around particle or grain boundaries (Fig. 13f). These features are especially notable on comparing the selected-area electron diffraction pattern inserts in Fig. 13e, f, respectively.

Figures 14 and 15 illustrate other examples of explosively consolidated metal powders in the context of their residual grain structures and microstructures, as well as the development of novel, residual properties as a consequence of the explosive

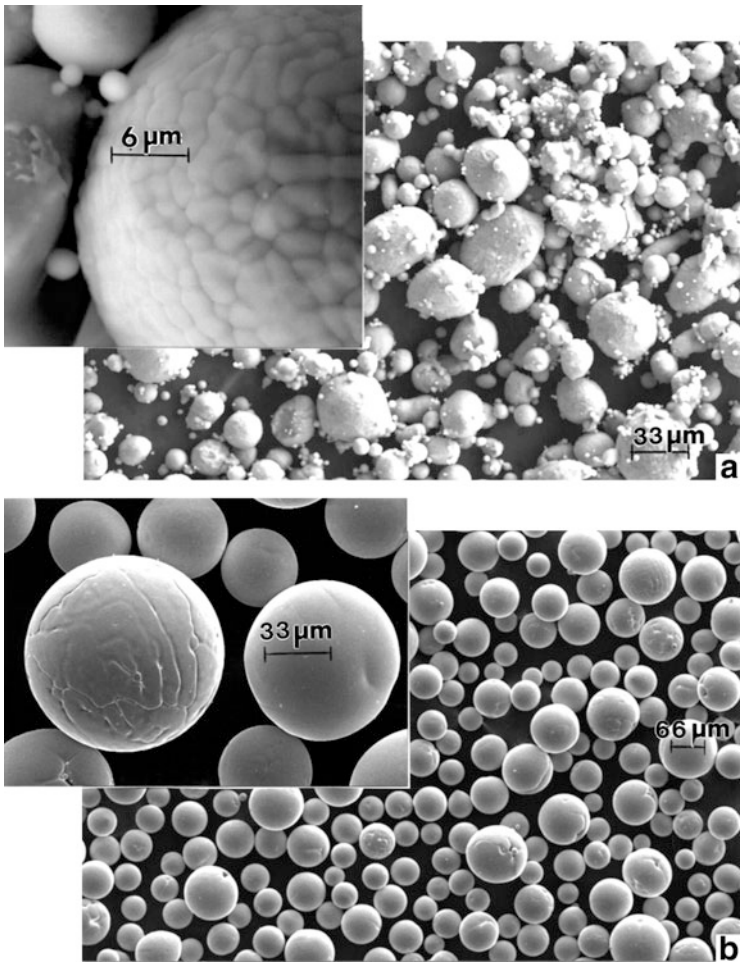


Fig. 12 Examples of rapidly solidified (atomized) metal and alloy powders, (**a**) Al-6061. *Insert* shows particle microdendritic structure. Satellite particles attached to larger particles provide small size distribution. Melting point ~650 °C. (**b**) Niobium. *Insert* shows particle microdendritic structure. Melting point ~2,500 °C

shock-wave consolidation process. Figure 14a shows variations of DRX zones characterizing the grain boundaries or Cu particle consolidation regimes, while Fig. 14b shows a very high density of dislocations corresponding to the grain interiors. Similarly, Fig. 15a shows dislocation of sub-grain microstructures in explosively consolidated molybdenum powder (having an initial powder size of ~5.4 μm). Correspondingly, Fig. 15b compares the initial, residual shock consolidation microindentation hardness (HV10) with more conventional Mo processed products and their recovery performance with increasing temperature.

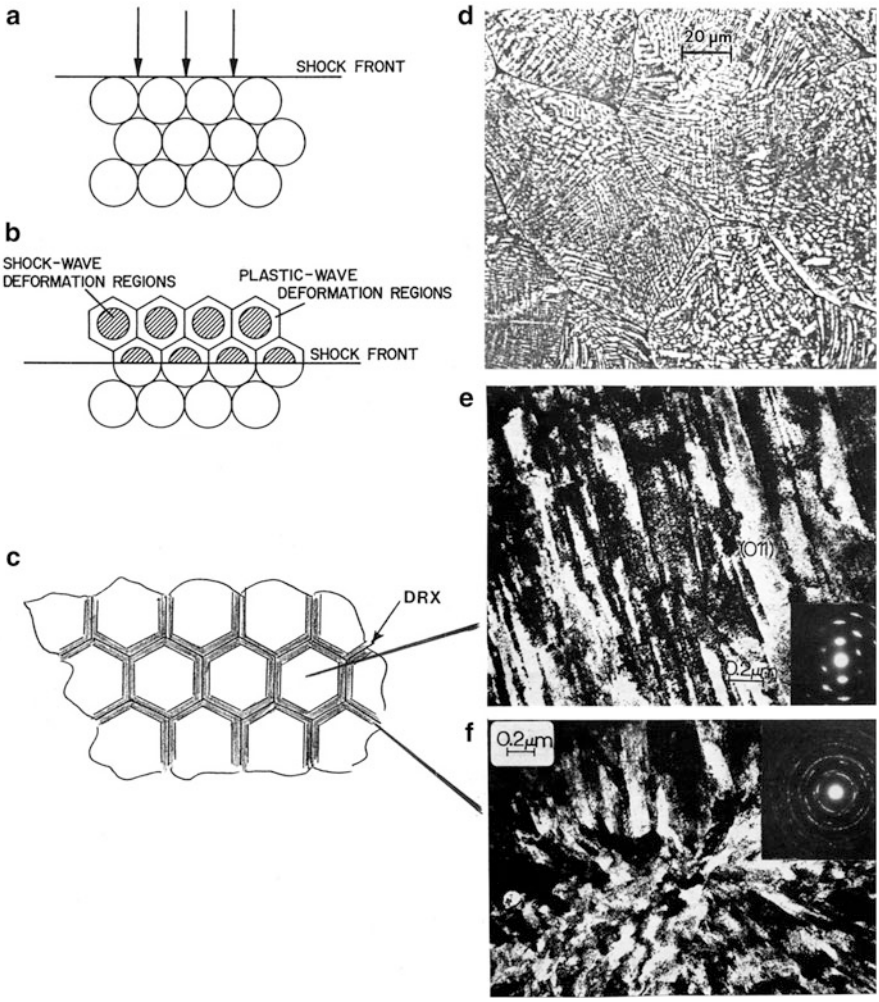


Fig. 13 Explosive consolidation of MAR-M200, Ni-base superalloy atomized powder using granular TNT. (**a**–**c**) show the powder consolidation by DRX production at their surfaces and deformation microstructures in particle interiors. (**d**) shows an optical micrograph of consolidated powder having microdendritic structure and equiaxed grain structure. (**e**, **f**) show TEM images for twin-related deformation structures in particle (grain) interiors and DRX fine-grained microstructures in the intergranular (interparticle consolidated) regions in (**c**) (Adapted from Meyers et al. (1988))

It is possible to fabricate consolidated tubes from metal and alloy powders by explosive consolidation by altering the configuration in Fig. 9 to that shown in Fig. 16, which was developed by Murr and Hare (1984) for the fabrication of bulk, amorphous tubes of magnetic, rapidly solidified amorphous ribbon-flakes (powder platelets) which could be used in a more convenient, conformal geometry than ribbon-cut plates to manufacture efficient transformers. ▶ Figure 3

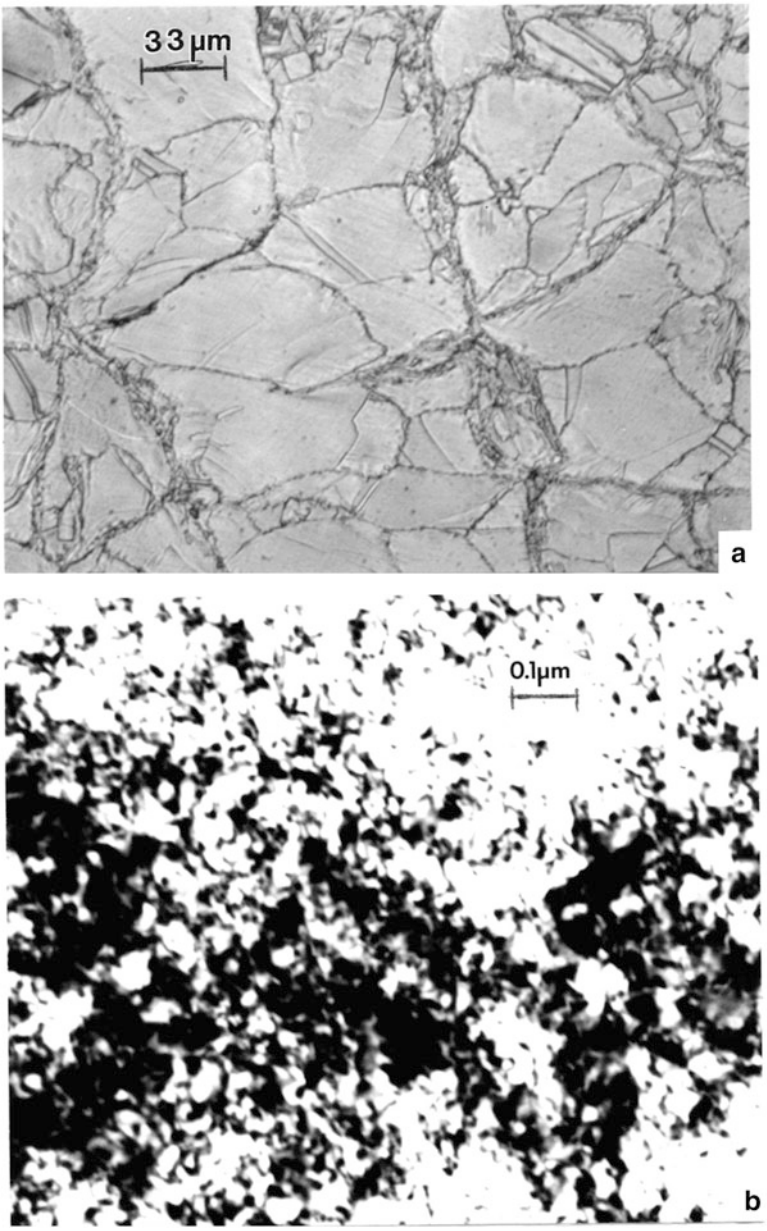


Fig. 14 Explosively consolidated Cu powder. (**a**) Optical micrograph showing grain (consolidated particle) structure containing a few annealing twins and fine DRX grains composing the interfacial/interparticle regime. (**b**) TEM image showing dense dislocation structure in Cu grain in (**a**). Explosive was ANFO: $V_D = 1.6$ km/s

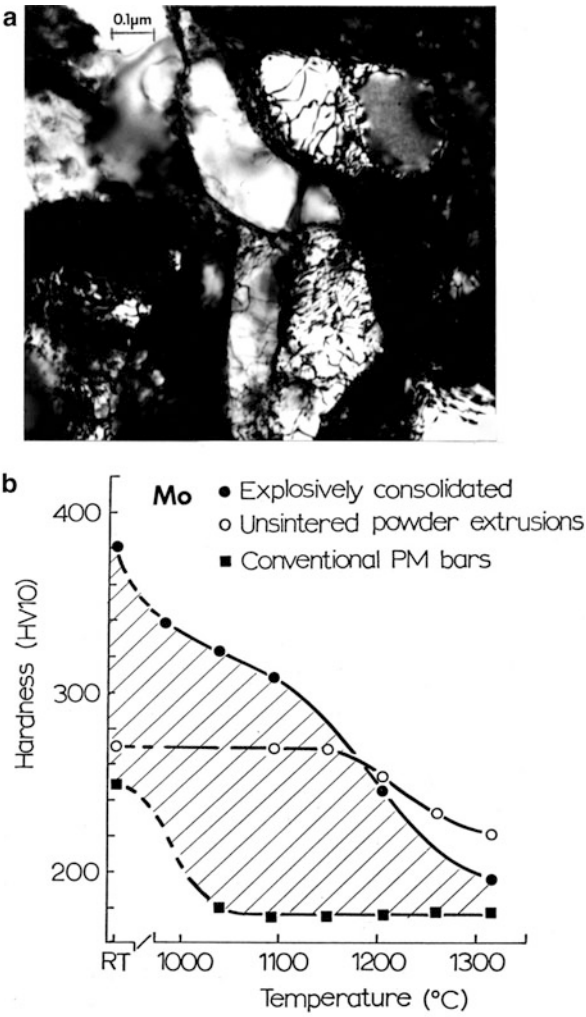


Fig. 15 Explosively consolidated Mo with 5.4 μm average powder size. (**a**) TEM showing nano-grain and dislocation structure for as-fabricated cylinder. (**b**) Microindentation hardness variation comparison for annealed Mo products indicated. Explosive was Amatol: $V_D = 4$ km/s

(chapter "Synthesis and Processing of Nanomaterials") illustrates schematically the supercooling of a melt stream to produce amorphous alloy ribbon which can be crushed to produce a precursor powder. Figure 17 illustrates an Fe-20B amorphous powder flake precursor and the resulting explosively consolidated product using the configuration shown in Fig. 16. Although TEM specimens prepared from the consolidated product in Fig. 17b did not illustrate any structure or microstructure, the corresponding selected-area electron diffraction pattern example shown as an insert in Fig. 17b demonstrates the amorphous structure is retained. Figure 17b also demonstrates the glass-like, conchoidal fracture surface

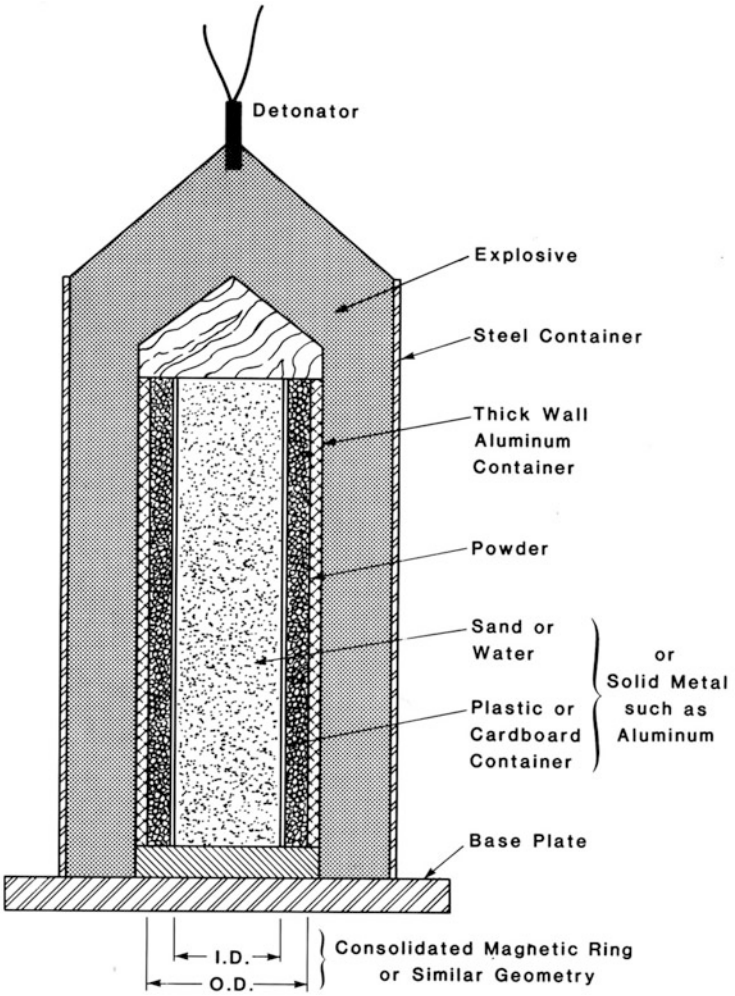


Fig. 16 Explosive consolidation fixture for fabricating hollow tube products such as magnetic ring precursors described by Murr and Hare (1984)

features for the amorphous, consolidated Fe-20B, indicative of the popular reference to amorphous metals and alloys as "metal glasses." One notable feature represented by Fig. 17b is the prospects for solid-state flow in the amorphous state without recrystallization, since there is no such evidence as demonstrated by the electron diffraction pattern insert. Figure 18 shows a modified version of the explosive consolidation configuration shown in Fig. 17 which allows for the production of individual, tubular or ring-type products. Of course the configurations shown in Figs. 16 and 18 are amenable to fabricating other explosively consolidated products as well.

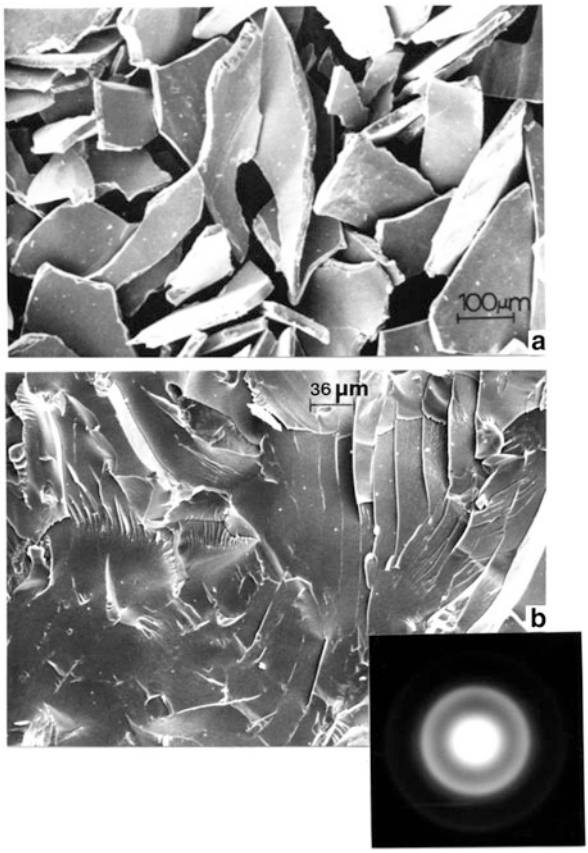


Fig. 17 Explosive consolidation of magnetic, amorphous, rapidly solidified Fe-20B powder platelets using the fixture of Fig. 16 and ANFO explosive ($V_D = 1.7$ km/s). (**a**) Initial powder prior to compaction. (**b**) Explosively consolidated powder platelets in (**a**). Fracture surface image in the SEM. *Insert* shows diffuse electron diffraction pattern for thin section prepared for TEM analysis confirming amorphous structure is retained after explosive consolidation

It was noted previously that powder properties such as hardness and melting point can have a significant influence on the optimization of explosive consolidation parameters in fabricating a good product. Figure 19 shows a general trend of shock pressure versus powder hardness ranging from aluminum and aluminum alloys (Fig. 12a) to diamond. This includes a wide range of powder particle sizes and size distributions as well as powder morphologies. It is apparent that in the larger, shaded portion of this trend line, which characterizes high-temperature ceramic powder materials, pressures exceed 50 GPa for consolidation. Consequently, explosive consolidation configurations would require very large dimensions and/or explosives (large detonation velocities, V_D). The flyer velocity can also be increased by a factor of 2 or more using a standoff or gap arrangement between the flyer tube and the powder tube as described by Meyers and Wang (1988) and illustrated in Fig. 20. For this explosive consolidation configuration, the flyer-tube velocity is given by

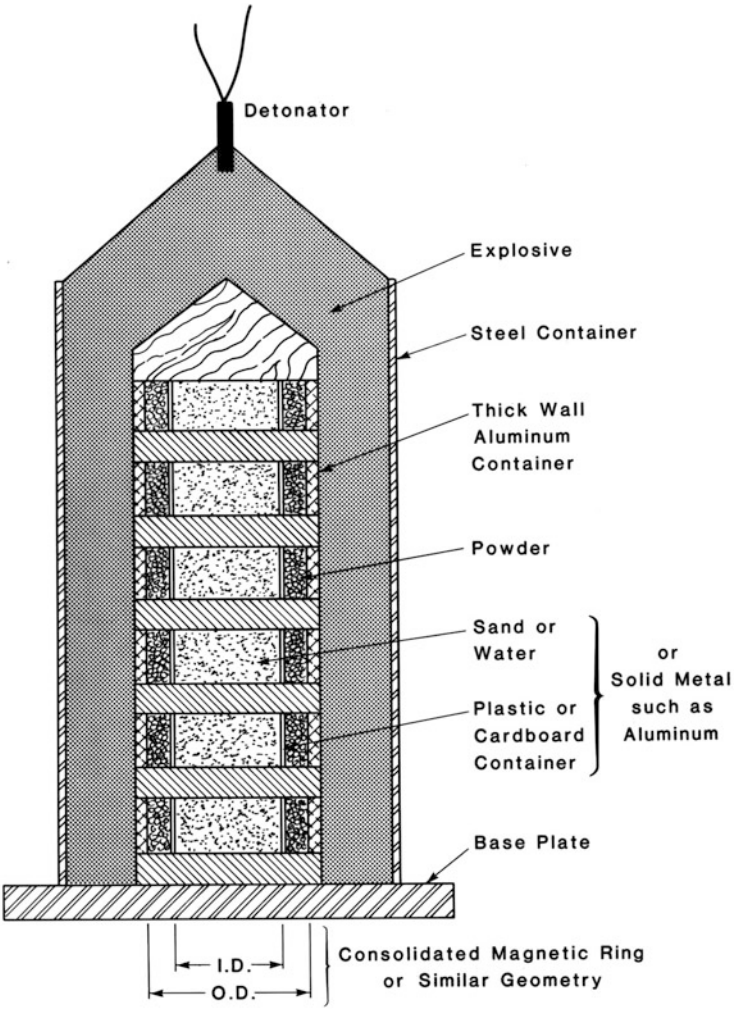


Fig. 18 Alternative explosive consolidation fixture (in contrast to Fig. 16) for producing segmented rings for magnetic winding applications of consolidated amorphous alloy products (Adapted from patent of Murr and Hare (1984))

$$V_f \cong \sqrt{2E}\left(3/\left[5\left(\frac{M}{C}\right)+2\left(\frac{M}{C}\right)\left(\frac{2R+r_o}{r_o}\right)+\left(\frac{2r_o}{R+r_o}\right)\right]\right)^{1/2} \tag{31}$$

where, as in Eq. 28, M/C = 1/R or

$$M/C = (\text{flyer tube volume})\rho_f/(\text{explosive volume})\rho_e \tag{32}$$

and R and r_o in Eq. 31 are the radii shown in Fig. 20. The standoff or tube gap shown is usually around twice the thickness of the flyer plate (h_f in Fig. 9). Meyers

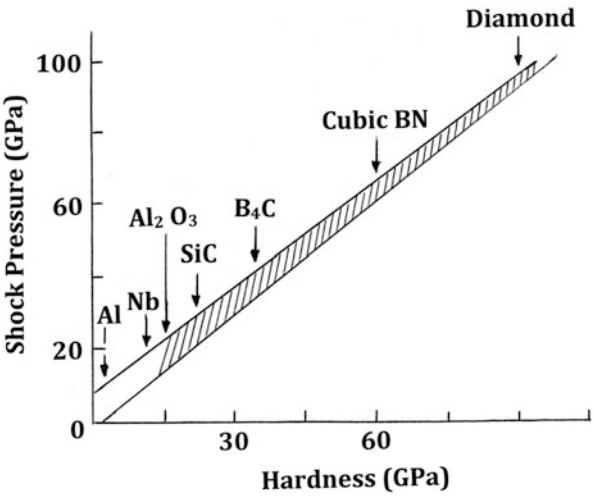


Fig. 19 Explosive shock pressure versus hardness for powder materials explosive consolidation. *Shading* represents hard ceramic materials (Adapted from data in Meyers et al. (1988))

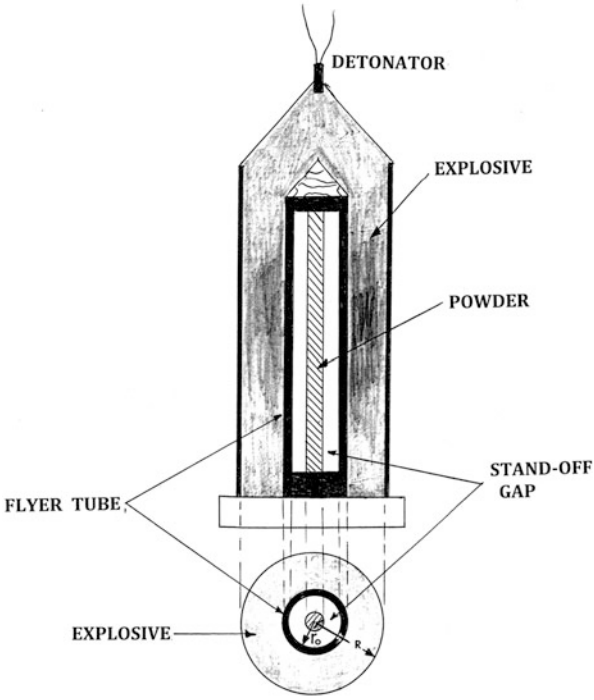


Fig. 20 Double flyer-tube fixture for explosive consolidation of powder in cylindrical configuration. The initial flyer tube is stood off from the metal powder container which acts as the second (contact) flyer. The gap is ~ twice the flyer-tube thickness. The outer container tube can be metal or another material. The flyer and powder container can be steel

et al. (1988) plotted the variation of $V_f/\sqrt{2E}$ with M/C shown reproduced in Fig. 21 for the implosion configuration illustrated in Fig. 20 in comparison with that for an expanding or exploding cylinder (as in Fig. 8a) and the open-faced sandwich configuration which characterizes explosive welding as shown in Fig. 1a. The hatching between curve A and B in Fig. 21 is indicative of the close

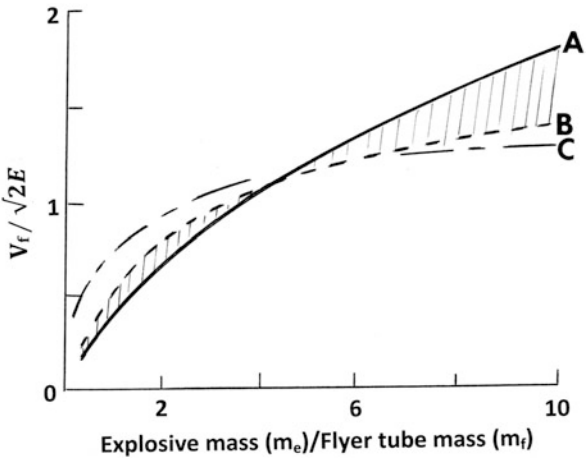


Fig. 21 Plot of $V_f/\sqrt{2E}$ versus C/M for explosive configurations. (**a**) Imploding cylinder in explosive consolidation (Figs. 9, 20). (**b**) Planar open-faced sandwich; explosive welding (Fig. 1). (**c**) Exploding cylinder, explosive forming (Fig. 8a) (Adapted from Meyers et al. (1988))

approximation between the implosion shock consolidation configuration and process description and the open-faced, explosive welding, flyer-plate process.

It is somewhat implicit from Fig. 21 that the explosive welding (sandwich regime) and the cylindrical, explosive consolidation regime are closely related (shading connecting A and B). Consequently, it is also possible to consolidate powders in sandwich (flat-plate tooling) arrays. These are described in the work of Meyers et al. (1988), Roman and Gorobtsov (1988), and others in the same reference source (Murr 1988). Novel configurations combining both explosive shock-wave consolidation of powders and explosive welding have also been developed by Murr (1991) for high-temperature superconductor products and conformal geometries.

Murr and Staudhammer (1988) have also reviewed shock-wave sensitization and reactivity of powdered materials where shock compaction induces high densities of dislocations to raise the volume free energy conducive to enhance chemical reactivity and new materials synthesis. This can apply to metals and alloys, polymeric powders, and ceramic powder materials or complex mixtures. In addition, explosive consolidation can provide a means to fabricate novel composites or composite mixtures of micron-size metal or alloy powders with nano-size ceramic powders or nanotubes and fibers especially in systems where there is no adequate wetting afforded by the molten metal matrix, e.g., SiC nanoparticles or fibers in Al melt (Alba-Baena et al. 2008).

References

Alba-Baena NG, Salas W, Murr LE (2008) Characterization of micro and nano second phase/2-phase regimes created by explosive shock-wave consolidation of powder mixtures. Mater Charact 59:1152–1160

Chadwick MD, Jackson PW (1983) Chapter 7: Explosive welding in planar geometries. In: Blazynski TZ (ed) Explosive welding, forming and compaction. Applied Science, London, pp 219–287

DeCarli PS, Jamieson JC (1961) Formation of diamond by explosive shock. Science 133:821–824
Deribas AA (1981) Chapter 52: Explosive metal working. In: Meyers MA, Murr LE (eds) Shock waves and high-strain-rate phenomena in metals. Plenum Press, New York, pp 915–939
Gurney RW (1947) Ballistic research report no 648, Aberdeen Proving Ground, Aberdeen
Meyers MA, Wang SL (1988) An improved method for shock consolidation of powders. Acta Metall 36(4):925–936
Meyers MA, Thadhani NN, Yu L-H (1988) Chapter: 8 Explosive shock wave consolidation of metal and ceramic powders. In: Murr LE (ed) Shock waves for industrial applications. Noyes, Park Ridge, pp 265–334
Murr LE (ed) (1988) Shock waves for industrial applications. Noyes, Park Ridge
Murr LE (1991) Explosive processing of bulk superconductors. Mater Manuf Process 6(1):1–32
Murr LE, Hare AW (1984) Implosive consolidation of a particle mass including amorphous material. US Patent 4,490,329, issued 25 Dec 1984. US Patent Office, Washington, DC
Murr LE, Hare AW (1988) Chapter 3: Explosive forming and material working applications. In: Murr LE (ed) Shock waves for industrial applications. Noyes, Park Ridge, pp 103–136
Murr LE, Staudhammer KP (1988) Chapter 12: Shock wave sensitization, shock-induced reactivity, and new materials fabrication. In: Murr LE (ed) Shock waves for industrial applications. Noyes, Park Ridge, pp 441–471
Murr LE, Hare AW, Eror NG (1988) Chapter 13: Fabrication of novel, bulk superconductor composites by simultaneous explosive consolidation and bonding. In: Murr LE (ed) Shock waves for industrial applications. Noyes, Park Ridge, pp 473–527
Rinehart J, Pearson J (1963) Explosive working of metals. MacMillan, New York
Roman OV, Gorobtsov VG (1988) Chapter 9: Shock wave powder compaction: state of the art and trends in the USSR. In: Murr LE (ed) Shock waves for industrial applications. Noyes, Park Ridge, pp 335–379
Salam SAL, Al-Hassani STS (1981) Chapter 57: Interfacial wave generation in explosive welding of multilaminates. In: Meyers MA, Murr LE (eds) Shock waves and high-strain-rate phenomena in metals. Plenum Press, New York, pp 1003–1018

Friction-Stir Welding and Processing

Contents

Abstract

Friction-stir welding (FSW) creates solid-state flow in butted metals and alloys which are mixed in the resulting weld zone. A rotating tool stirs one material into the other by the creation of DRX grains which result by extreme deformation, creating adiabatic heating at the tool/material interface. The process bonds or joins same or dissimilar metals and alloys and is not limited by thermodynamic considerations in conventional fusion welds. Unlike fusion welding, the tool is not purposely consumed. Friction-stir processing (FSP) uses a rotating tool as in FSW with some dimensional restrictions to allow controlled surface depths to be altered, microstructural alteration. Continuous tool traverse and overlapping stir regimes produce a modified or remodeled surface region or thickness where second phase as well as other microstructures can be homogeneously distributed. Bubbles or voids can be eliminated in cast materials. Properties and performance such as superplasticity can be affected by creating DRX nanograins.

Introduction: Friction-Stir Welding

Friction-stir welding (FSW) involves the solid-state joining of two butted or faying surfaces using a rotating tool as illustrated schematically in Fig. 1. The tool itself is not purposely consumed in the process as in conventional fusion welding, and there is no melting, and the weld zone represents the flow of one material into the other. In the context of extreme deformation, it is aphoristic to state that this process

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_51

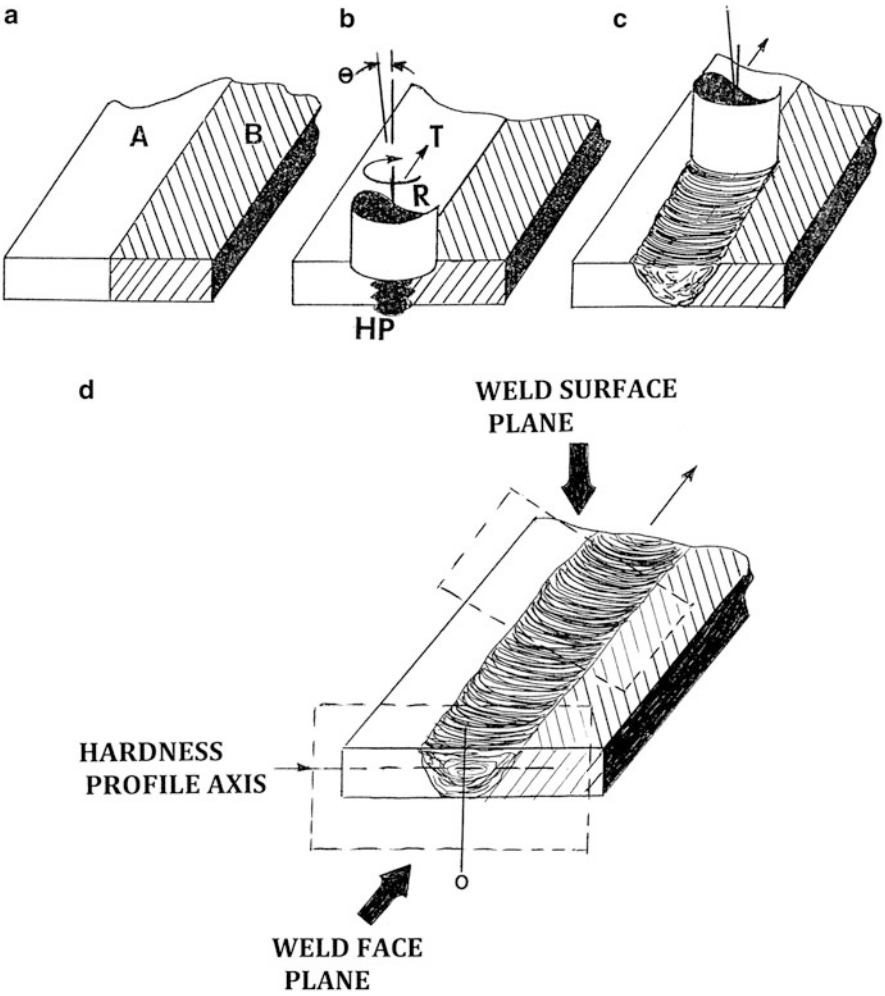


Fig. 1 Friction-stir welding (*FSW*) schematic views. (**a**) Butted or faying plates (<1 mm to >10 cm). *A* and *B* can be the same or dissimilar materials. Same plate compositions can also involve dissimilar microstructures and properties, for example, significant hardness differences, etc. (**b**) Rotating tool or pin tool or head pin (*HP*) begins traverse (*T*) along faying surfaces. Different rotation speeds (*R*) and traverse speeds can optimize any particular FSW process. The tool rotation axis is usually a few degrees from the normal axis (θ). The tool holder or shoulder is forced against the surface plane which also influences the weld profile as illustrated in (**c**). Clockwise (cw) rotation of the tool as in (**b**) characterizes plate A as the advancing side while *B* is the retreating (or trailing) side of the weld. (**d**) shows the completed weld and provides a reference system for observing the weld face (weld-face plane) and the weld surface (weld-surface plane). HV denotes the convention for measuring microindentation hardness profiles through the weld zone

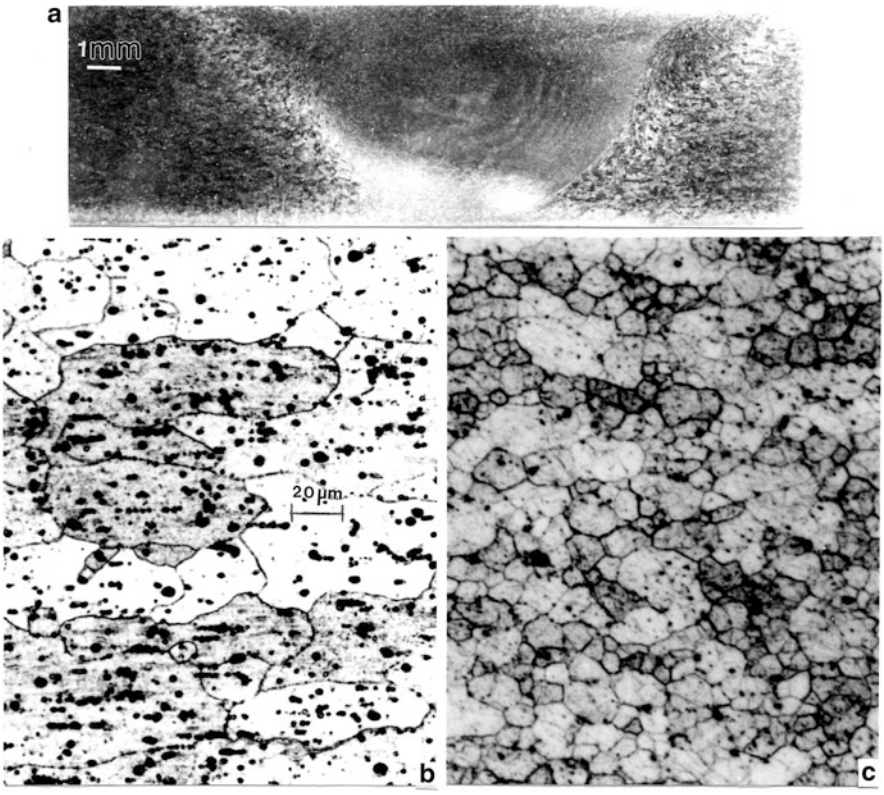


Fig. 2 Weld-face view of the weld zone through FSW plates of Al-6061 alloy 0.63 cm thick. (**a**) Weld with onion ring (flow) structure. The right side is the advancing side (tool rotating counter-clockwise (ccw) at 400 rpm (R = 400 rpm ccw); T = 2 mm/s. (**b**) Base grain structure. (**c**) Weld-zone center grain structure at same magnification shown in (**b**)

epitomizes large-scale dynamic recrystallization (DRX). This is demonstrated on comparing typical examples of FSW of Al-6061, a common, lightweight alloy having numerous aerospace applications where FSW has been applied since 1991 when Thomas et al. (1991) at The Welding Institute (in England) invented the process.

Figures 2–4 show typical FSW microstructures for Al-6061 alloy from the perspective illustrated in the schematic views of Fig. 1. In Fig. 2, the solid-state flow in the weld zone is apparent in the so-called onion structure, more apparent in the left portion of the weld zone in Fig. 2a. It can also be observed in Fig. 2a that the right side of the weld zone has a more well-defined demarcation or interfacial regime than the more diffused zone at the left. This phenomenon is characteristic of the stirring of one side of the fayed plates into the other which can be conceptually

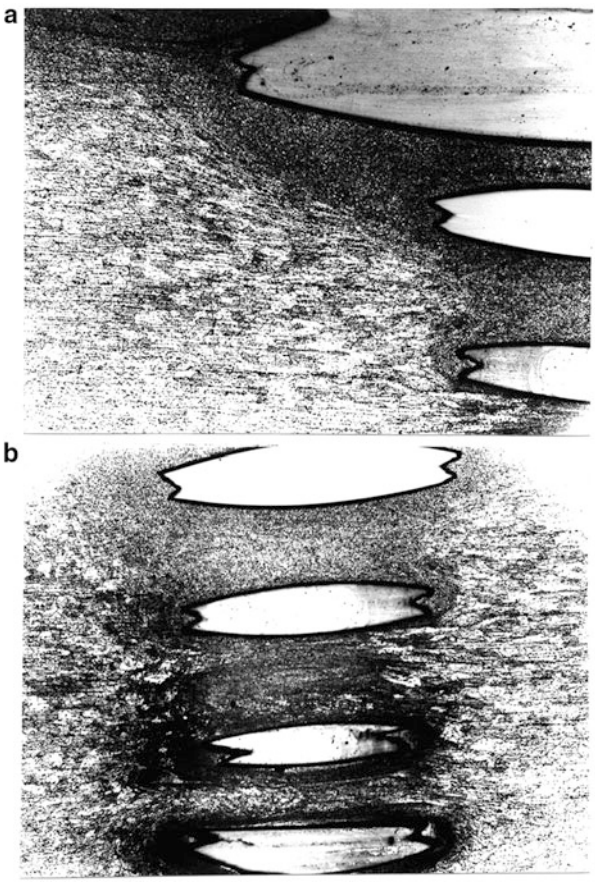


Fig. 3 Cross section showing hardened steel screw head pin tool in an Al-6061 FSW sample as in Fig. 2. (**a**, **b**) shows offset views from the top of the screw

described by the flow of one side, the advancing side, into the other, creating a retreating side. The base material grain structure shown in Fig. 2b is refined in the weld zone as shown in Fig. 2c, where the DRX grain size is increased by grain growth as the tool advances or traverses along the adjacent plates as shown by T in Fig. 1b. This weld-zone grain growth occurs because the temperature in the initial weld zone formation at the head of the advancing tool reaches roughly 0.8 T_M, where T_M is the melting temperature of the welded plates. Figure 3 shows these features in a cross section of the "frozen," advancing tool, which in this particular case is a simple hardened steel screw mounted in a milling head. Figure 4a shows a color rendering of the base material grain structure converging to the weld zone in friction-stir welded Al-6061. Faint stir pattern (onion rings) can be observed in the weld zone at DRX, which also characterizes the weld-zone grain structure. The interfacial regime between the weld zone and the base material is also observed to exhibit deformation and distortion of the base material grain structure, and this zone along with the actual DRX weld zone is often referred to as the stir-affected zone

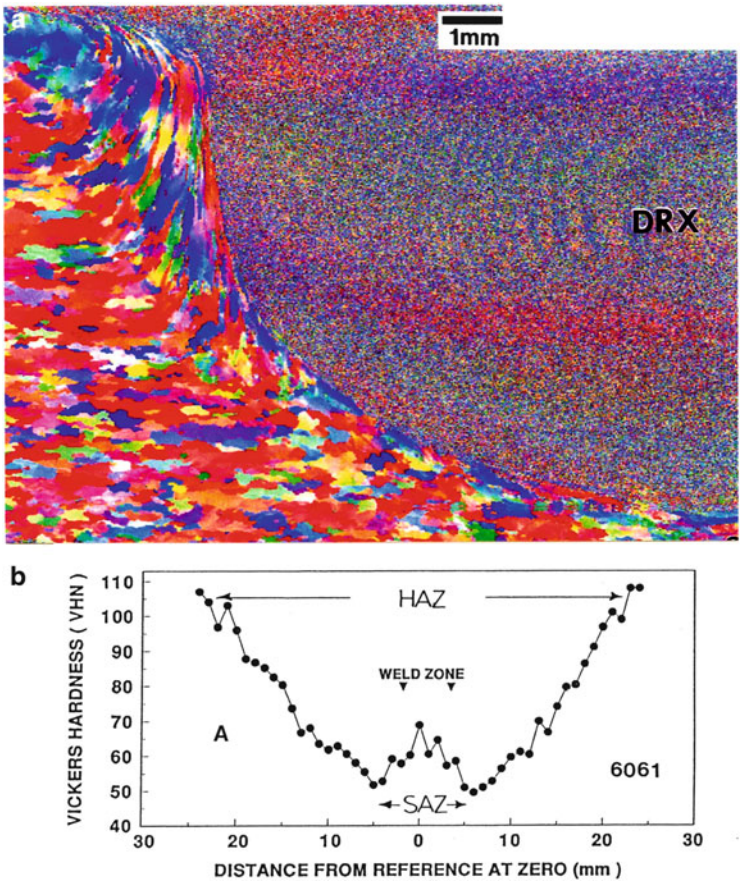


Fig. 4 (**a**) Electron backscatter diffraction (*EBSD*) grain structure map in the SEM showing an FSW zone half section in Al-6061. *Colors* denote common grain orientations which appear to texture along the weld-zone interface. Note onion structure to the left of DRX. *Color difference* for these striations indicates microstructure variations. (**b**) Weld-zone microindentation hardness profile for an Al-6061 FSW zone similar to (**a**): *A* to *left* shows advancing size ((**a**) is courtesy 9f Dr. John McClure, UTEP)

(SAZ). However, as in conventional fusion welding, the heat generated in the weld zone extends beyond the actual DRX-FSW zone and beyond the bounds of the SAZ, or thermomechanically affected zone (TMAZ). This heat-affected zone (HAZ) also includes the weld and the SAZ and has an influence on the annealing effects on microstructure within this regime, including of course the DRX grain structure (size).

Figure 4b illustrates the microindentation hardness measured for friction-stir welded Al-6061 alloy as in Figs. 2–4a from the weld center into the base material. Here, the SAZ and HAZ can be clearly delineated with respect to residual microstructures illustrated in Figs. 2–4a. The average reduction in microindentation hardness in the FSW zone in contrast to the base material in Fig. 4b is observed

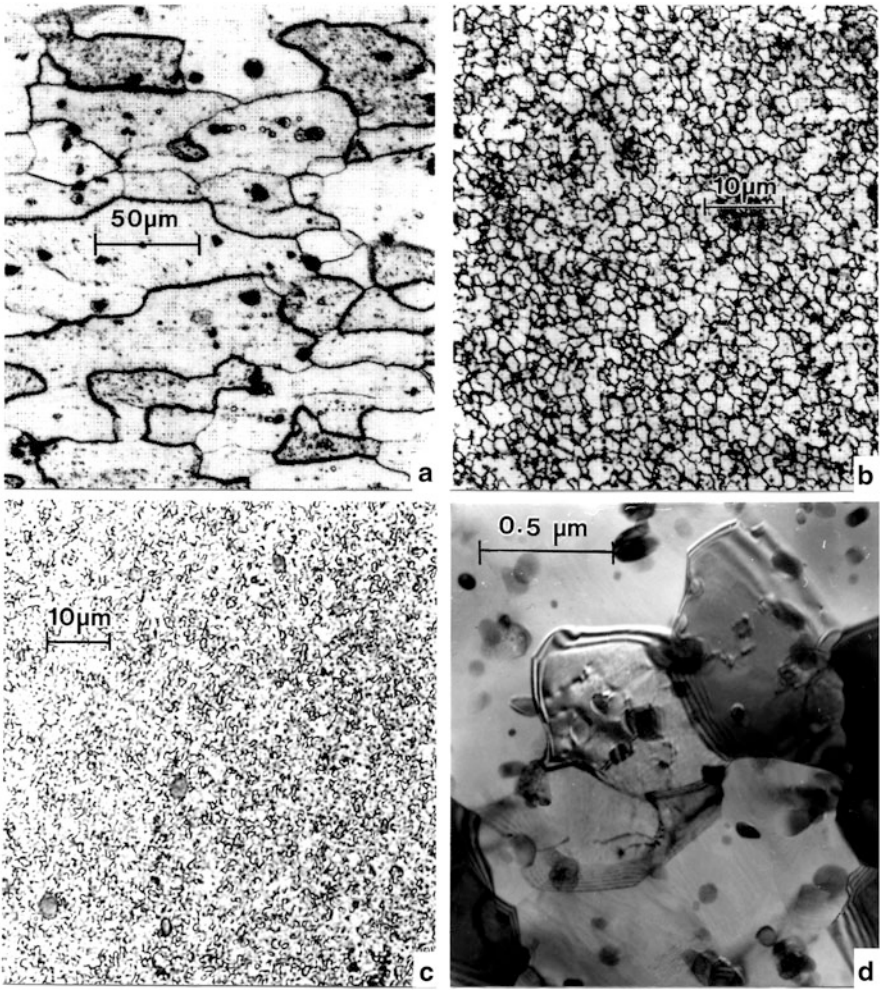


Fig. 5 Comparison of base material and weld-zone grain structures for Al-2024 alloy FSW at "room temperature weld start" (**a**, **b**) and for the workpiece cooled to −100 °C (**c**, **d**). (**a**, **b**) shows the workpiece (base) and FSW zone structure. (**c**) is the FSW zone grain structure. (**d**) shows a TEM image of (**c**). (Adapted from Murr (2010))

to be ∼45 % as a consequence of the dissolution of Widmanstätten (precipitate) structures in the weld zone.

In alloys not so much affected by precipitation annealing such as Al-2024, the degradation of hardness is not as drastic. In addition, if the grain growth can be suppressed by cooling the FSW workpiece, the microindentation hardness can also be better controlled. These features are illustrated in the microstructure observations in Fig. 5 and corresponding microindentation hardness profile in Fig. 6 for FSW of Al-2024 alloy. Although the microstructure difference between the base

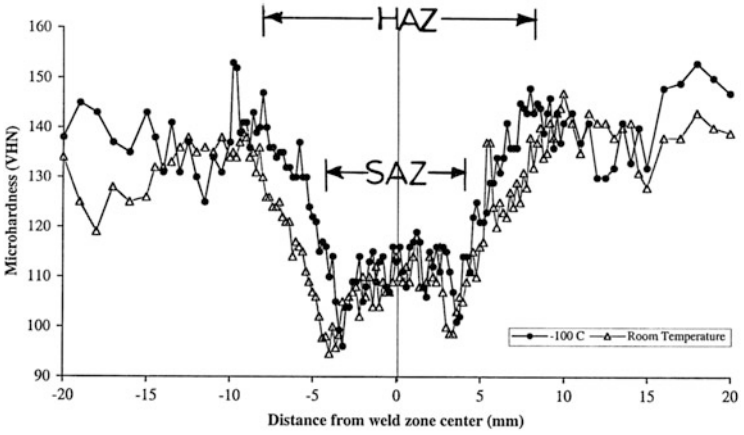


Fig. 6 Comparison of microindentation hardness profiles for "room temperature" FSW start and workpiece cooled to −100 °C (under liquid nitrogen) for Al-2024 alloy in Fig. 5. Stir-affected zone (*SAZ*) and equivalent heat-affected zone (*HAZ*) are indicated. HAZ is also referred to as a thermomechanically affected zone (*TMAZ*) in FSW. (After Benevidez et al. (2000))

material and the FSW zone for the Al-2024 shown in Fig. 5a, b is essentially the same as that shown for Al-6061 alloy in Fig. 2b, c, there is a notable difference between the base material and the FSW zone grain structure (grain size) for Al-2024 when the workpiece is cooled to −100 °C using liquid nitrogen, as illustrated on comparing Fig. 6a, c. As shown in the corresponding TEM image for the FSW weld zone in the Al-2024 (Fig. 5d), the equiaxed grain size is ∼500 nm in contrast to ∼2 μm in the uncooled weld zone represented by Fig. 5b. The corresponding comparison for the microindentation hardness profiles shown in Fig. 6 indicates a noticeable difference in the hardness variation between the cooled and uncooled workpiece. In addition, the average hardness reduction in the weld zone (SAZ) for FSW of Al-2024 is observed to be ∼20 % in contrast to ∼45 % for FSW of Al-6061 shown in Fig. 4b.

In contrast to the microindentation hardness reduction in the weld zone for the FSW of Al-6061 and Al-2024 shown in Figs. 4b and 6, respectively, the FSW of a beryllium-aluminum composite (62 Be-38 Al: Be spheroids in an Al matrix) results in a Be-Al nanograin, mechanically alloyed weld zone as a result of DRX and little or no grain growth and a corresponding residual hardening in the weld zone (SAZ). This is illustrated in Figs. 7 and 8. The SAZ hardness increase from the base material is shown in Fig. 7b to be ∼40 % in contrast to the ∼45 % decrease for Al-6061 FSW shown in Fig. 4b.

Figures 9 and 10 provide another interesting perspective for FSW. In Fig. 9a, 20 vol% Al_2O_3 particles are homogeneously dispersed in an Al-6061 equiaxed, polycrystalline matrix. Following FSW, these dispersed particles remain homogeneously dispersed (Fig. 9b), but the Al-6061 grain size is reduced from the base material in a manner similar to the FSW zone reduction shown for Al-6061 FSW in Fig. 2b, c and the Al-2024 FSW in Fig. 5a, b. Figure 10 compares the

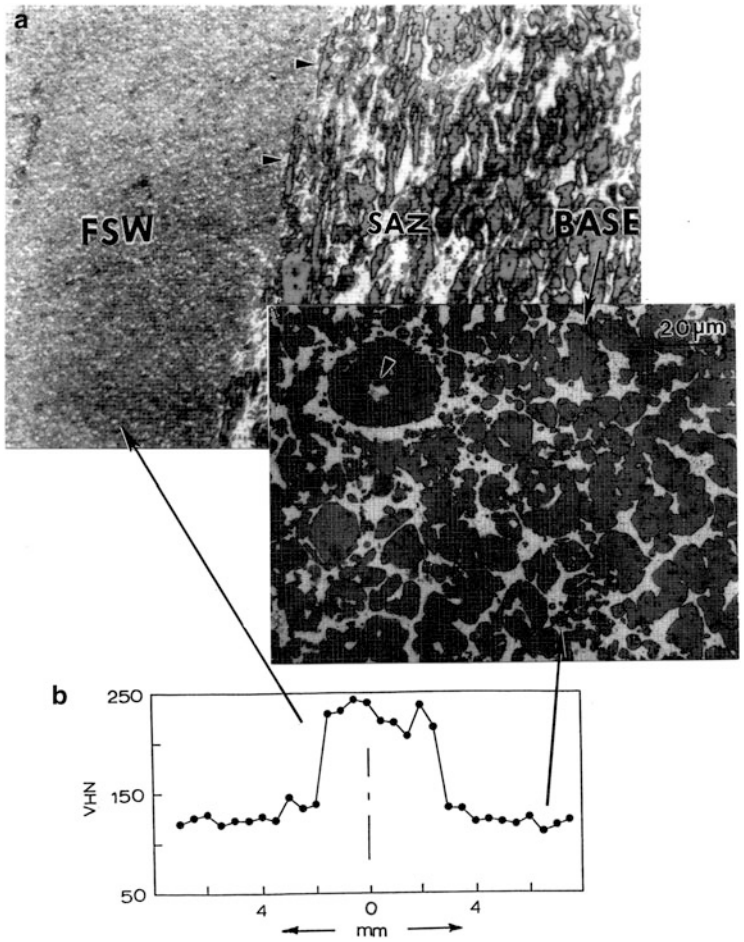


Fig. 7 FSW of AlBeMet (powder metallurgy) alloy (62Be-38Al) composite: Be particles in an Al matrix. (**a**) Weld zone-(*FSW*)-SAZ-Base material are indicated. *Insert* shows details of the composite base structure (*arrow* shows occluded Al in Be particle). (**b**) Microindentation hardness profile through the weld zone represented in (**a**). FSW parameters: R = 1,000 rpm, ccw; T = 1 mm/sec. (Adapted from Contreras et al. (2001))

microindentation hardness profiles across the weld center in Fig. 9c with a corresponding FSW zone center for Al-6061 similar to that shown in Fig. 4a. The hardness reduction in the FSW zone for the Al-6061/Al_2O_3 in Fig. 10 relative to the base material hardness is ~13 % in contrast to ~33 % for the undispersed Al-6061. Of course it is possible to re-heat-treat FSW workpieces in some configurations to re-establish requisite microstructures (usually precipitates) in order to recover the microindentation hardness loss.

The solid-state flow by DRX grain boundary sliding which facilitates the FSW process and the creation of a weld zone is well documented in the examples

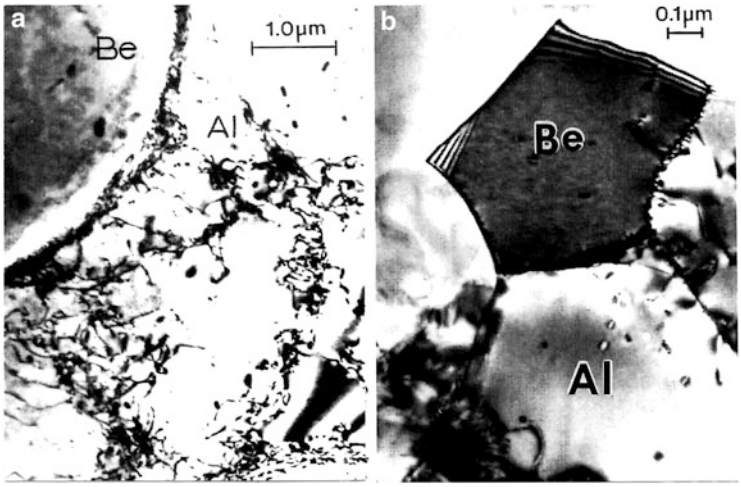


Fig. 8 TEM images corresponding to the base (**a**) and FSW zone (**b**) microstructures in the 62Be-38Al powder metallurgy alloy in Fig. 7**a**). (After Contreras et al. (2001))

provided in Figs. 2, 3, 4a, 5, 7, 8, and 9. Figure 9 provides a particularly interesting flow concept. It is certainly apparent that in order to homogeneously mix the Al_2O_3 particles in the two fayed plates of Al-6061/Al_2O_3 as shown in Fig. 9a, b, requires these particles to flow in the Al-6061 DRX matrix. The ability of solid particles or even unrecrystallized blocks of material to be transported (or flow) in a fluid-like, solid-state regime is also demonstrated in the FSW example provided in Fig. 11. In Fig. 11a a top view looking down into the advancing FSW pin tool (at 90° to the view provided in Fig. 3 showing the tool cross section) is shown. The pin tool, rotating counterclockwise (ccw), is advancing into Al-6061 plates having a thin Cu strip ~0.2 mm thick inserted between the butted plates. The rotating (R = 1,000 rpm) tool deforms and particulates the Cu insert at P in Fig. 11a, and these particulated Cu particles flow around the tool and into the weld zone forming behind the tool. The Al-6061 workpiece plates are recrystallized by the advancing tool, forming a solid-state flow regime which transports the Cu particles as shown in Fig. 11b. This flow regime (the Al-6061 matrix DRX regime) can form DRX layers having varying thicknesses, which can flow one over the other. Figure 11c shows this feature schematically, while Fig. 12 shows a more systematic view of the formation of the FSW zone by DRX and subsequent grain growth due to residual annealing by adiabatic heat retention within the weld zone. The formation of DRX bands as well as variations in FSW-induced deformation within these bands implicit in the DRX-related microstructure evolution in Fig. 12 also accounts for the banding characterizing the onion structures shown in the weld-face views of Figs. 2a, 4a, and 9c, in particular, and as a consequence of etching variations producing contrast in these bands. The bands are also characteristic of the solid-state flow or flow patterns which arise by tool rotation and translation.

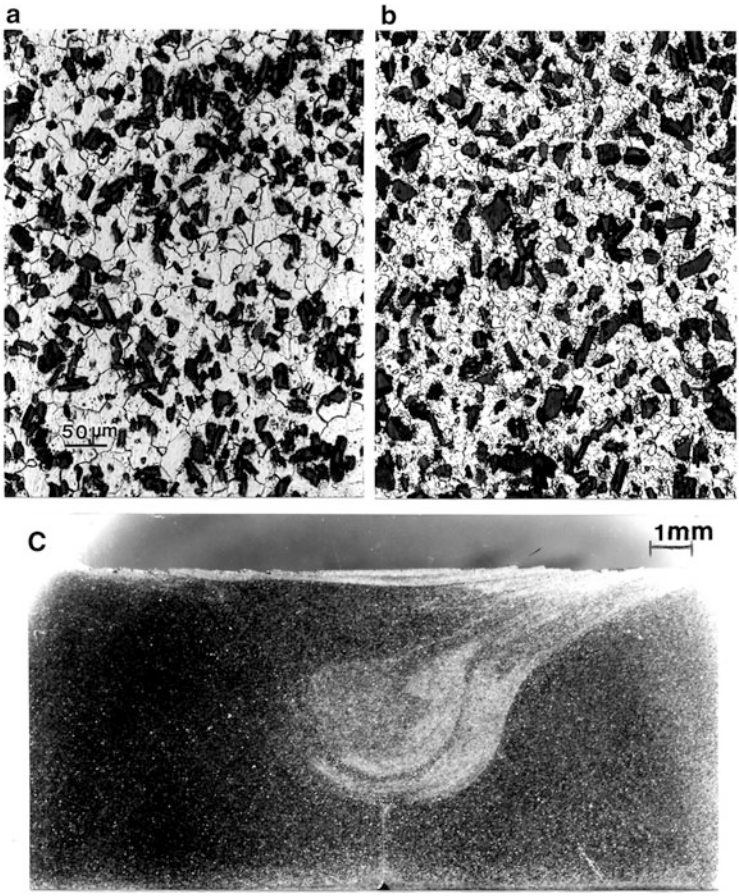


Fig. 9 FSW of Al-6061/20 vol.% Al_2O_3. (**a**) Base (unwelded plate) microstructure. Al_2O_3 particles are black dispersion. (**b**) FSW zone microstructure. (**c**) Weld-face view of FSW section. Weld parameters: R = 650 rpm; T = 1 mm/s. Magnification of (**a**) and (**b**) is the same as shown in (**a**)

It might also be noticed in examining Fig. 11a that the strain associated with the deformation of the Cu insert can be measured directly by comparing the initial Cu thickness shown, with the thickness it is stretched to just before or at the point of particulation (P in Fig. 11a). This results in a true strain in excess of 4. Correspondingly, the strain rate can be estimated from the tool rotation (rotational displacement) and the rate of rotation (1,000 rpm). The strain rate is estimated to be $>10^3\ s^{-1}$. Consequently, the process conditions are conducive to adiabatic heating and DRX production.

It should also be noted that the tool holder, referred to as the tool shoulder, has a significant effect on the weld formation since the weld zone is shaped from the bottom of the rotating tool to the shoulder dimension as shown schematically in Fig. 1c and in the actual weld-zone profiles shown in Figs. 2a, 4a, and 9c. To a large extent, the force of the shoulder against the workpiece will influence the actual

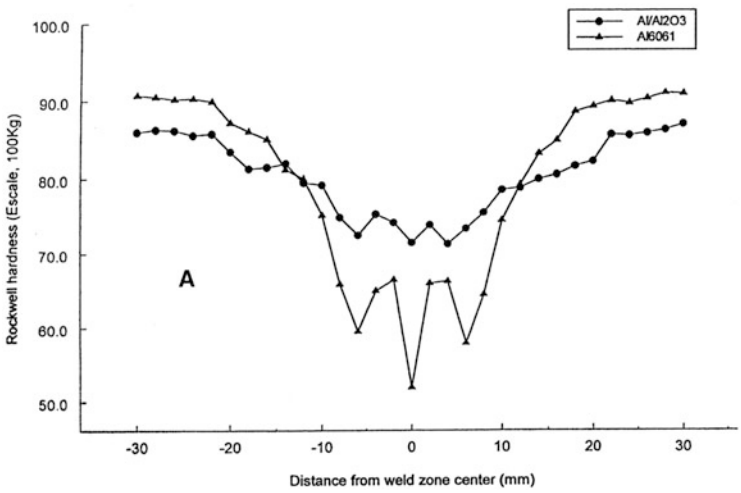


Fig. 10 Comparison of FSW microindentation profiles for Al-6061 and Al-6061/20 % Al_2O_3 (Fig. 9). *A* indicates the advancing side. Weld parameters are the same: R = 650 rpm; T = 1 mm/s

stirring by the shoulder where both the shoulder and the tool are involved in the DRX formation at the surface and along the fayed surfaces which are stirred by the pin tool. It is apparent from Fig. 11a that tool rotation will influence the strain rate: increasing rotation, R, will increase the strain rate, ˙. In addition, the translation, T (Fig. 1b), of the tool and shoulder regime will also significantly influence the FSW process, in some respects influencing the flow stress associated with the process. In this respect, the solid-state flow (fluid-like flow) can be related to an effective viscosity (μ) (Colegrove and Shercliff 2005):

$$\mu = \sigma/3\dot{\epsilon}, \tag{1}$$

where the σ is the flow stress, and the strain rate, $\dot{\epsilon}$, can be expressed generally by the *Zener-Holloman equation* originally developed for the hot deformation of metals in the form:

$$\dot{\epsilon} = A_z \sinh\left(\frac{\sigma}{\sigma_R}\right)^n \exp\left(\frac{-Q}{RT}\right), \tag{2}$$

where the parameters A_Z, σ_R, Q (the activation energy), and n are obtained by fitting the equation to an experimentally measured flow stress for a given strain ($\in$) and temperature, T. R in Eq. 2 is the gas constant.

While the heat generated in the FSW process can be considered to be primarily adiabatic and related to the product, $\varepsilon \times \dot{\epsilon}$, some frictional heating can also arise especially at the shoulder and related to the pressure applied to the shoulder, the rotation speed (or angular velocity), and some friction coefficient. In some respects, the process is dominated more by adiabatic than frictional phenomena, and friction-stirring is a misnomer.

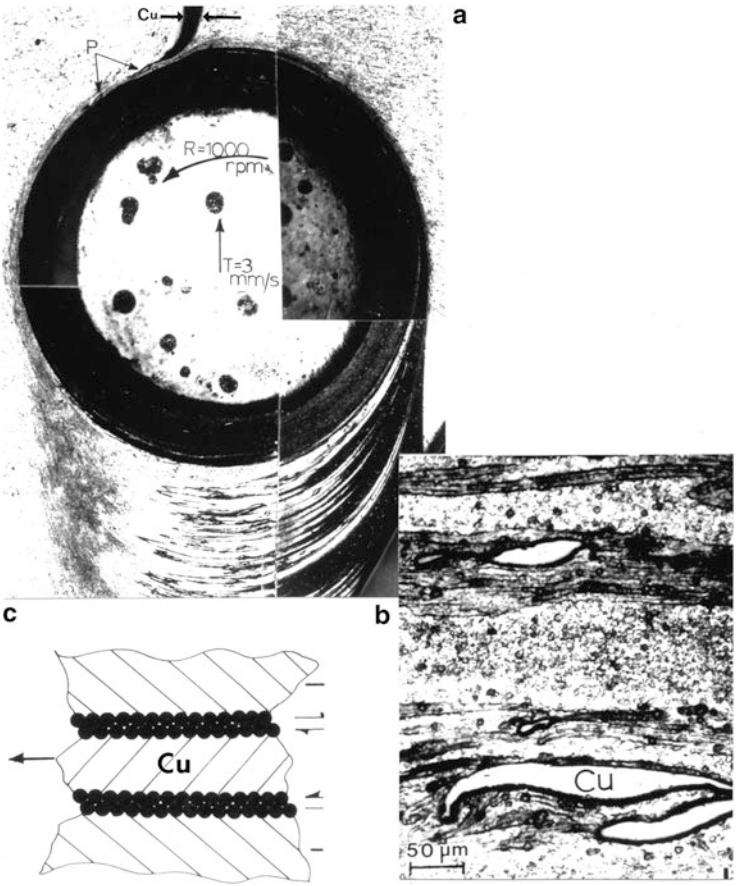


Fig. 11 FSW of Al-6061 plates with a copper ribbon insert. (**a**) View of the pin tool in a surface plane section (Fig. 1**d**). Pin tool traverse is to the top. Cu particulates can be observed in the weld zone. (**b**) shows a magnified view of the weld zone with Cu particles flowing in the Al-6061 alloy DRX matrix. (**c**) Schematic view of Cu particle flowing between thin DRX lamellar

Friction-Stir Welding of Dissimilar Metals and Alloys

Like explosive welding, friction-stir welding can avoid the thermodynamic limitations of fusion joining or bonding by creating a DRX solid-state flow regime which can allow for mixing and bonding through complex flow regimes. These are somewhat apparent in the residual flow patterns illustrated in Figs. 2a, 4a, 9c, and 11c in particular for the FSW of the same material in butted plate configurations. The joining of dissimilar metals and alloys by FSW poses a particularly novel and useful joining technology especially in joining conventionally unweldable configurations. In addition, from a microstructural perspective, dissimilar system FSW allows for

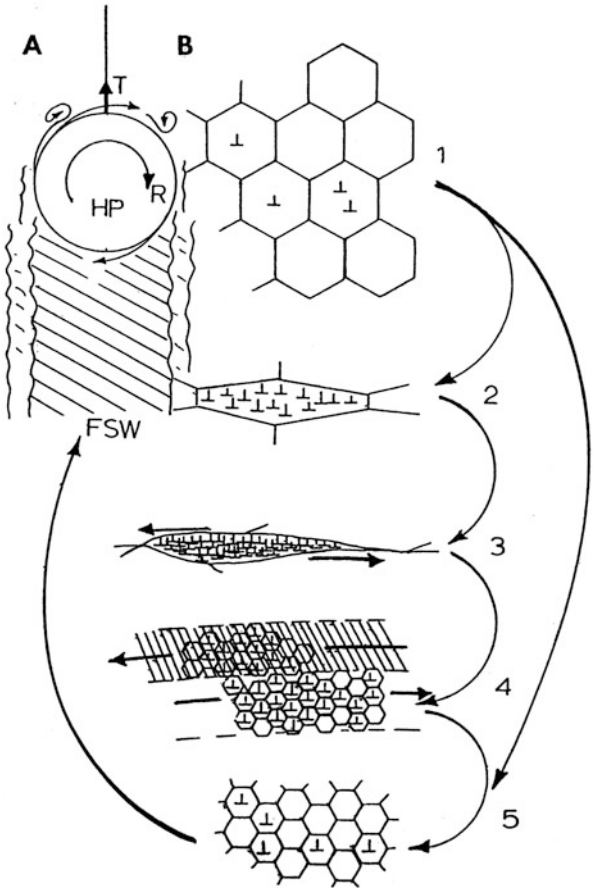


Fig. 12 FSW process schematic from the pin tool perspective. (*1–2*) grain deformation, (2–*3*) severe deformation and high dislocation production, (*3–4*) DRX, (*4–5*) grain growth. Actual process involves complex DRX flow, vortex formation, DRX lamellar intercalation, and block flow of unrecyrstallized segments

more distinct visualization of complex flow patterns because of the differences in composition and microstructures which etch differently and often promote more distinct microstructural contrast when viewed by light metallography or scanning electron microscopy.

Figure 13a shows the weld face and weld zone for FSW of Al-2024 (Al-4.5 Cu, 0.5 Mg, 0.3 Mn) as the advancing side plate to Al-7039 (Al-4 Zn, 2.8 Mg, 0.3 Mn, 0.2 Cr) as the retreating side (tool in Fig. 1b rotating clockwise (cw) as shown). This produces a sharp interface at the weld zone on the left (advancing side) in contrast to a diffuse interfacial transition on the right side (retreating side) of the weld zone in Fig. 13a. This is similar to, but reversed from, the weld-zone structure shown for Al-6061 FSW in Fig. 2a. However, the flow pattern (onion structure) is much more pronounced in Fig. 13a in contrast to Fig. 2a. Figure 13b illustrates a more extensive weld-zone microstructure section cut from an Al-2024/Al-6061 (Al-1 Mg, 0.65Si, 0.25 Cu, 0.2 Cr) dissimilar FSW system. The complexity of the flow pattern is illustrated along the weld as it is associated with the weld-face flow as well as flow

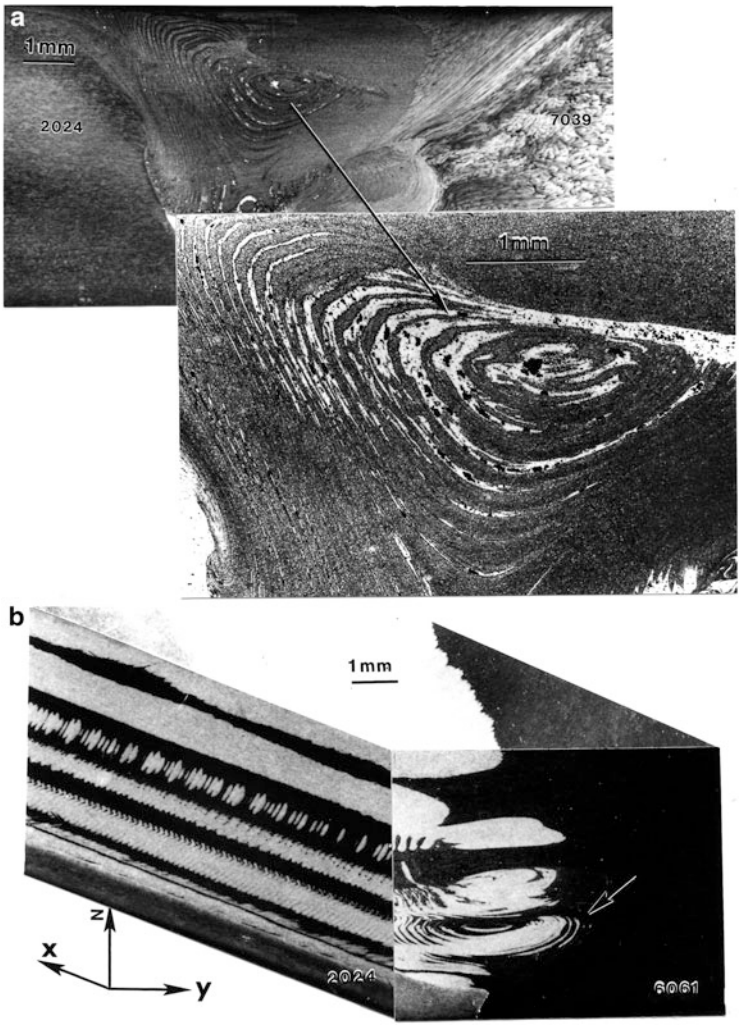


Fig. 13 Complex, solid-state flow patterns characterizing the weld-face onion structure (**a**) and the weld-center structure in a 3D section (**b**) for the FSW of Al-2024/Al-7039 dissimilar weld-plate system. (After Murr et al. (1999))

associated with the tool shoulder at the surface. Figure 14a and (b) illustrates simple flow phenomena associated with a screw-type head pin or tool (Fig. 1b), as well as a smooth tool configuration in Fig. 14a. In the configurations of Fig. 14, the tool rotates counterclockwise (ccw) and the right-side plate (Al-6061 in Fig. 13b) becomes the advancing side. Figure 14b shows schematically the geometry for the section view in Fig. 14b which is revered from the advancing tool configuration in Fig. 14a. It is apparent in Figs. 14b and 13b that the flow pattern within the weld length is related to the tool screw thread structure (Fig. 3). Correspondingly, the tool design can have an

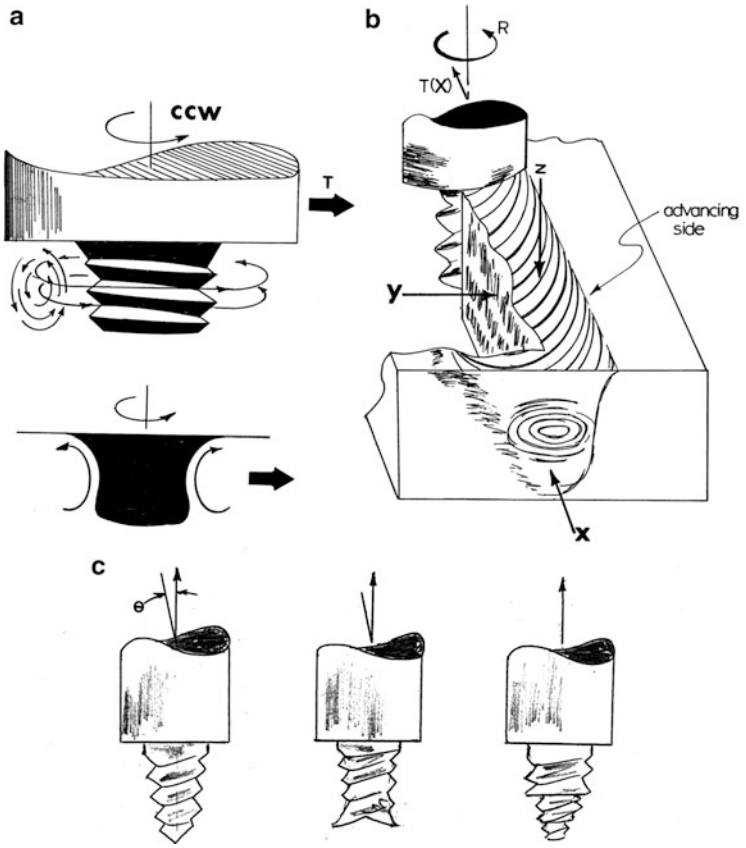


Fig. 14 Schematic views of solid-state flow lines and flow features associated with tool (and shoulder) rotation and stirring (**a**) and (**b**) tool traverse 3D section characteristic of Fig. 13b. (**a**) shows threaded and smooth fluted tool. (**c**) shows general tool concepts: tapered threads, flared geometry, and graded, thick section tool (*left to right*). Tool materials: Ni- and Co-base alloys, stainless steels, tool steels, Ti alloys, W alloys (especially for high-temperature materials FSW). Tool dimensions are also changed for various forms of FSW

effect on the weld efficiency and integrity, and Fig. 14c illustrates several commercial FSW tool designs. These designs are especially important for welding difficult-to-join materials such as metal matrix composites illustrated in Fig. 9 and other high-temperature metals and alloys such as stainless steel, Ti, and Ti-alloys, where in extended welding there may be tool erosion (or wear) and the tool temperature may be $\sim 0.8\ T_M$, which for high-temperature metals and alloys will cause softening as the tool advances over long weld lengths. The two pin tool views shown in Fig. 14a in fact represent the evolution of an equilibrium or optimized geometry or geometrical configuration with tool erosion which has been studied in the FSW of metal matrix composites such as Al-6061/Al_2O_3 (Fig. 9) and Al-339/SiC and related systems (Prado et al. 2003).

Figures 15 and 16 show numerous examples of complex flow patterns for a variety of dissimilar FSW systems as observed in the weld-face plane normal to the X-direction in Fig. 14b and the weld-surface plane normal to the Z-direction in Fig. 14b. It is especially notable to observe complex vortex mixing patterns in the weld-surface plane which are the result of the shoulder mixing in contrast to more regular flow of DRX zones and unrecrystallized, base material microstructure (grain structure) zones. Figure 15 illustrates these flow pattern variances for different, dissimilar FSW systems as noted, where Fig. 15b, c corresponds to the same dissimilar system: Cu/Ag. Figure 15a, b is in the weld face while c, d is in the surface. Figure 16 shows only comparative FSW weld-face plane views for DRX intercalation regimes which flow one over the other as shown schematically in Figs. 11c and 17. Figure 16c shows an SEM view which is complementary to the light microscope views in Fig. 16a, b for the Cu/Ag dissimilar weld. All of the weld-face plane views in Figs. 15 and 16 show the sharp interfacial advancing side of the weld zone.

Figure 18 compares residual microindentation hardness profiles across the weld zone for several dissimilar metal and alloy FSW systems. It is interesting to note the side-to-side trend variance of these dissimilar FSW systems in contrast to symmetrical side-to-side hardness profiles characteristic of same FSW systems shown in Figs. 4b, 6, and 10 where the base hardness is the same on both the advancing and retreating sides of the weld. Another feature of the weld-zone hardness profiles for the dissimilar FSW systems shown in Fig. 18 is the extreme hardness measurement variations which are skewed to one side of the weld in Fig. 18c, b and more symmetrically oriented around the weld center in Fig. 18c. These hardness variations correspond to the intercalations and flow features which can be skewed to one side or another of the weld center as shown in the weld-face view in Fig. 13a for the advancing side (Al-2024 (left) side). Shifting or reversing the side will essentially reverse the hardness profile. Of course changing the tool rotation direction changes the advancing and retreating side configurations which also changes the hardness profiles in the weld face. There are also variations in hardness profiles with tool rotation parameters, R and T (Fig. 1b), and it is easier to stir a very hard material into a softer material than the reverse. As illustrated in Fig. 15a, b, it is not necessary that both butted (dissimilar) materials to be welded by FSW dynamically recrystallize. In situations where the two components differ significantly in melting point, the lower-melting material will normally form the DRX regime (or flow matrix), although this is a somewhat more complex issue.

Over the past two decades, FSW has been widely adopted around the world in many engineering and manufacturing sectors especially aeronautical and aerospace, automotive, ship building, rail systems, and many others (Singh 2012). Many industries now manufacture large, commercial FSW systems. Large sections can be friction-stir welded up to 10 cm on a side, and double-sided welds can weld sections as thick as 0.2 m. Many welding configurations other than planar arrays can be welded, including various forms of spot welding, butt, lap, and related extrusion and laminate welding. As noted previously, special,

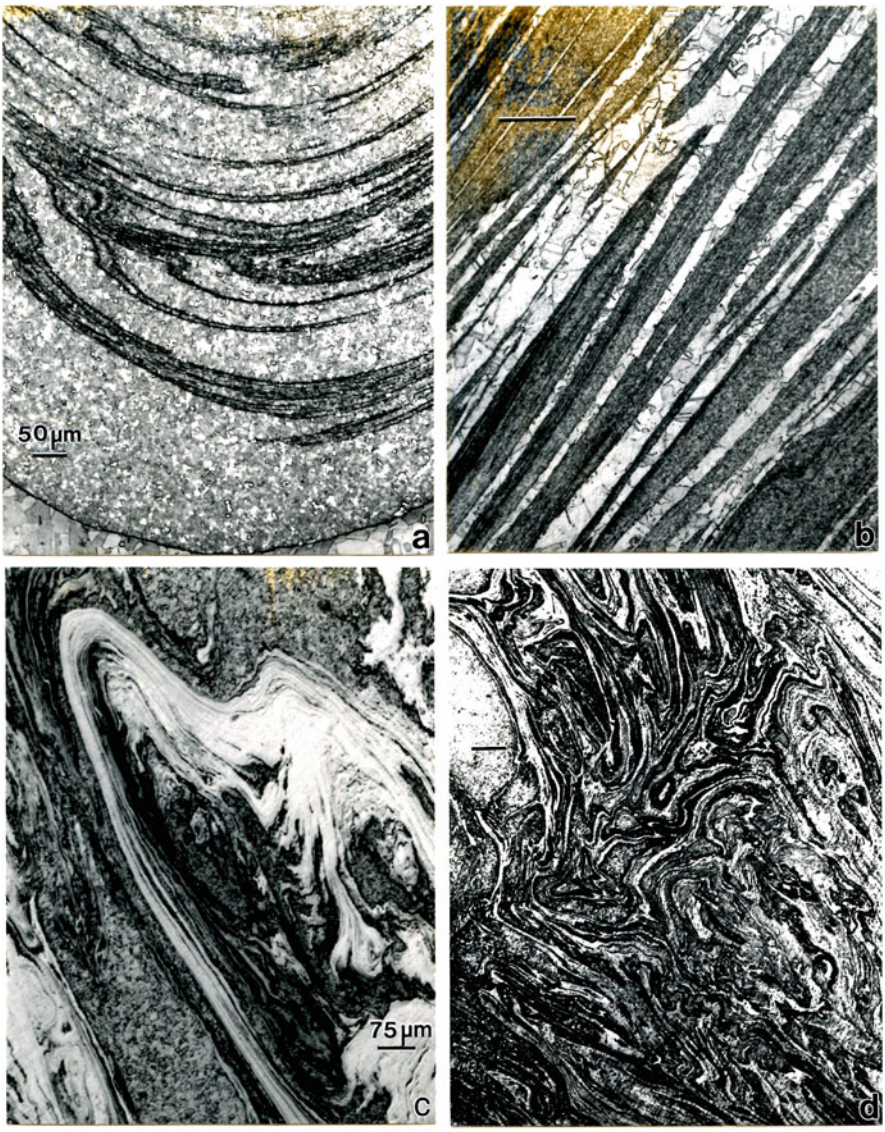


Fig. 15 Flow patterns characterizing the weld face and workpiece weld surface for FSW of dissimilar metals and alloys. (**a**) Cu/brass (70 Cu-30 Zn): weld-face plane (Fig. 1c). (**b**) Cu/Ag: weld-face plane. (**c**) Cu/Ag: weld-surface plane. (**d**) Brass/Ag: weld-surface plane. All FSW parameters: R = 1,000 rpm, T = 1 mm/s (From Murr (2010)). Magnification maker in (**b**) corresponds to (**a**). Marker in (**d**) corresponds to (**c**)

high-temperature ceramic tools such as WC allow for high-temperature metal and alloy welding, particularly in smaller thicknesses where adequate heat dissipation can be accommodated. Many other adaptations of friction welding have been used commercially for decades. These include solid and hollow pipe welding where

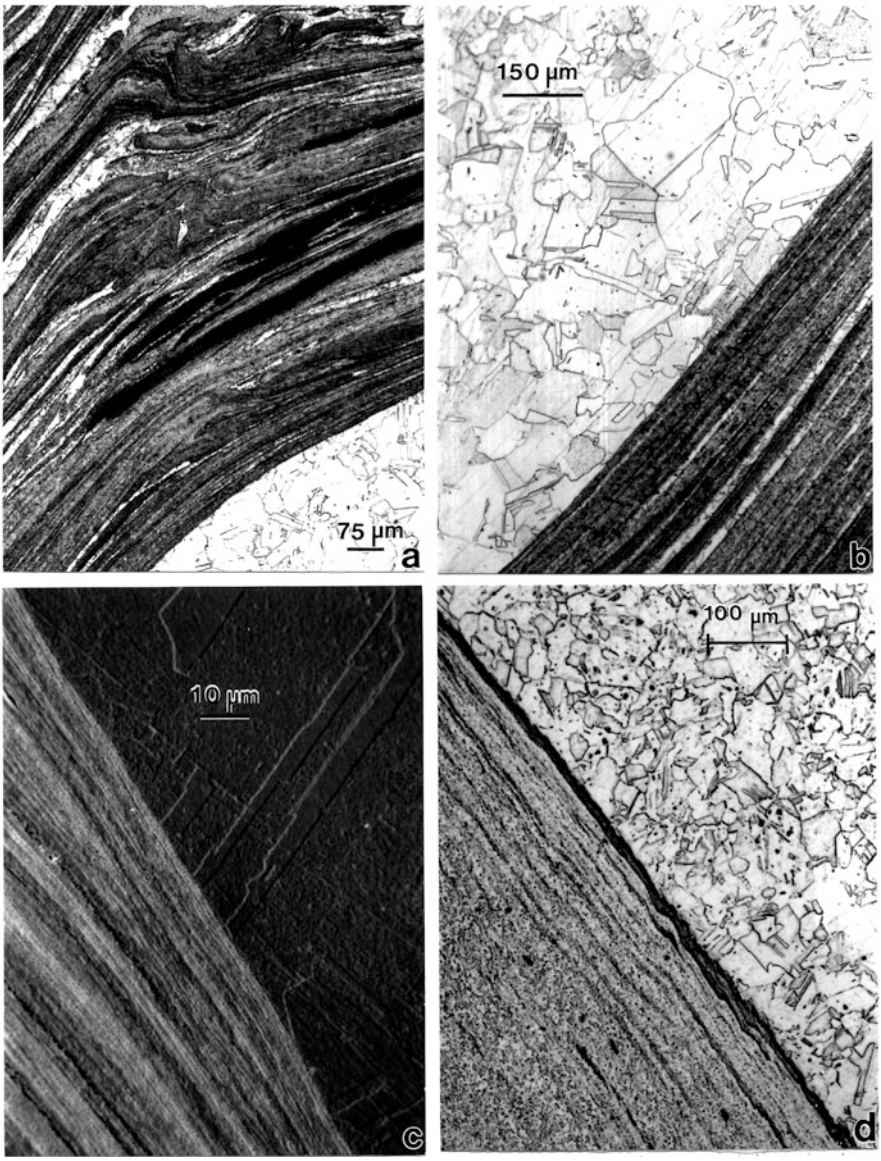


Fig. 16 Flow patterns and DRX intercalation patterns for dissimilar FSW systems. All views are in the weld-face plane (Fig. 1c). (**a**, **b**) shows different views in a Cu/Ag weld. (**c**) SEM view of Cu/Ag FSW zone. (**d**) Brass/Ag FSW zone (From Murr (2010))

one member is rotated and forced against the other, and the friction rotation welding of pipe sections to flat platforms and the like. This includes welding of different pipe and related geometry materials, etc., as illustrated herein for planar (metal and alloy) configurations.

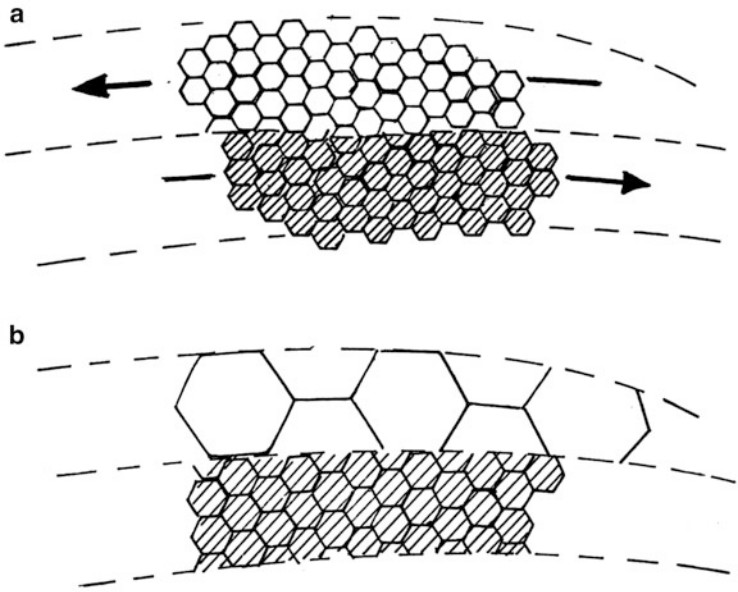


Fig. 17 Schematic views for DRX intercalation lamellae. (**a**) Shaded band is deformed relative to undeformed, DRX band. (**b**) Unrecrystallized grain structure flowing over deformed DRX band

Friction-Stir Processing

As illustrated previously in ▶ Fig. 11 of chapter "Synthesis and Processing of Nanomaterials" friction-stir processing (FSP), a slight modification of FSW, can be used to create nanocrystalline regimes in a variety of processed metals and alloys in particular. On a somewhat broader basis, FSP can be used to alter materials properties in the near surface or deeper regimes for forged, extruded, rolled, or cast metals or alloys or metal or alloy composites, such as *metal matrix composites MMCs* (Arora et al. 2012). FSP can be especially useful in modifying the structure and microstructure of cast metals or alloys because it can homogenize the microstructure to a depth of the tool. This can include the elimination of porosity in cast materials (Sun and Apelian 2011). Nonhomogeneous second-phase inclusions can be homogeneously distributed in a DRX-grown grain structure as implicit in Fig. 9 for an Al-6061/20 % Al_2O_3 MMC. Softer phase inclusions can also be broken up and more homogeneously, mechanically mixed by FSP processing.

Figure 19a, b reproduces ▶ Fig. 11 of chapter "Synthesis and Processing of Nanomaterials" as a convenience in viewing various FSP concepts. Figure 19c illustrates the actual modification of a surface section on a material by multiple head pin (tool) passes which are illustrated to have a three forward and three reverse traverse sequence. This can be altered in various ways to achieve the desired microstructural modification. In addition, multiple cycles can be performed to

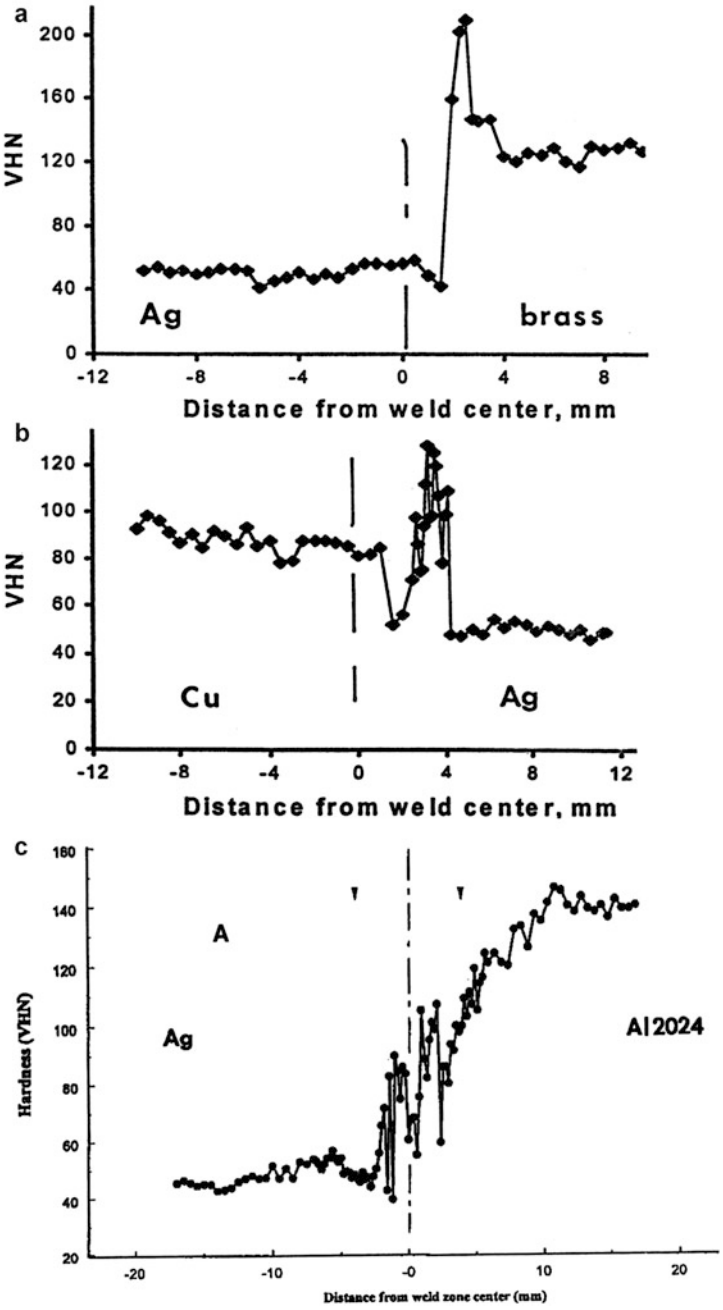


Fig. 18 Weld (*FSW*)-zone/weld-face residual (Vickers) microindentation hardness profiles for dissimilar systems noted. The advancing side is noted *A* in (**c**). *Arrows* in (**c**) show mid-weld dimension. (From Murr (2010))

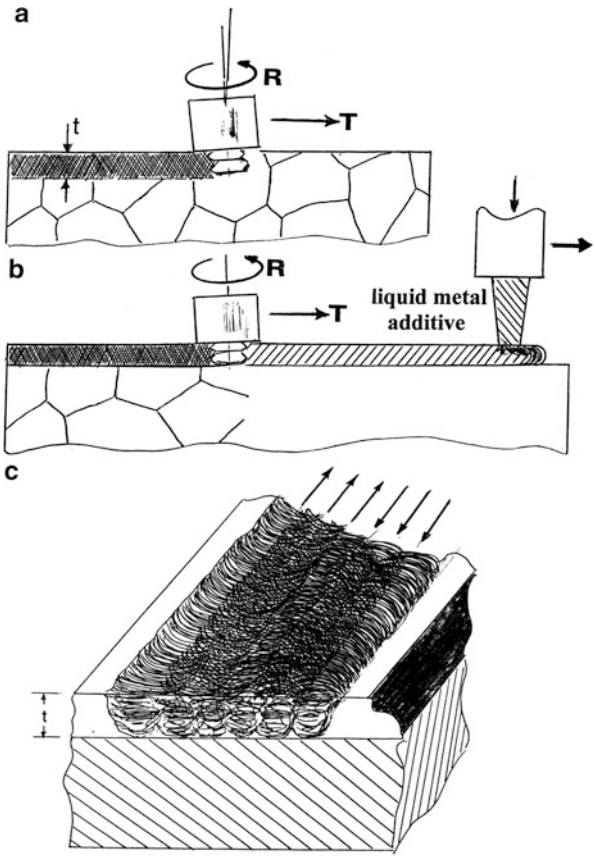


Fig. 19 FSP schematic views. (**a**) FSP of a plate-surface regime of thickness, t. (**b**) FSP of coated surface application. A liquid (melt) bead is applied and cooled prior to FSP modification. This layer can also be applied by spray coating or related processes. (**c**) Continuous FSP processing of a surface area. Traverse directions (schedules or sequencing) are illustrated

assure more homogeneity or to affect additional or alternate microstructure variations. For example, in Fig. 19c the traverse spacing can change to allow varying overlap of the stirred zone, and repeated cycles may then fill in (or more homogeneously mix) unstirred regions especially at the bottom of the tool. It is apparent from Fig. 19a, b that the depth of the FSP zone (t) can be altered by changing the tool length. Since the FSP zone is often same fraction of the workpiece to be altered, the shoulder or tool holder surface design can also be modified to help facilitate either DRX formation or the more homogeneous mixing of the flowing material. The workpiece can also be cooled to achieve a finer, nanograin size corresponding to the intrinsic DRX grain structure as implicit on comparing Fig. 5b, c.

There are many additional thermomechanical alterations of the FSP regimes which can be developed as shown schematically in Fig. 19a, b, including post- or pre-rolling of the workpiece, pre- and post-heat treatment to manipulate (age or re-age) microstructures, etc. This regime can also be subjected to shock-wave hardening using a layer of explosive detonated on the processed surface, along with other treatment schedules to assure specific or novel properties and performance.

References

Arora HS, Singh H, Dhindawl BK (2012) Composite fabrication using friction stir processing – a review. Int J Adv Manuf Technol 61(9–12):1043–1055

Benevidez S, Li Y, Murr LE (2000) Ultrafine grain structure in the friction-stir welding of aluminum alloy 2024 at low temperature. In: Mishra RS, Semiatin SL, Suryanarayana C, Thadhani NN, Lowe TC (eds) Ultrafine grained materials. The Minerals, Metals and Materials Society, Warrendale, pp 155–163

Colegrove PA, Shercliff HR (2005) 3-Dimensional CFD modeling of flow round a threaded friction stir welding tool profile. J Mater Process Technol 169:320–327

Contreras F, Trillo EA, Murr LE (2001) Friction-stir welding of beryllium-aluminum powder metallurgy alloy. J Mater Sci 37(1):89–99

Murr LE (2010) A review of FSW research on dissimilar metal and alloy systems. J Mater Eng Perform 19(8):1017–1089

Murr LE, Trillo EA, Li Y, Flores RD, Nowak BM, McClure JC (1999) Solid-state flow associated with the friction-stir welding of dissimilar metals. In: El-Kaddah N, Robertson DGC, Johansen ST, Voller VR (eds) Fluid flow phenomena in metals processing. The Minerals, Metals and Materials Society, Warrendale

Prado RA, Murr LE, Soto KF, McClure JC (2003) Self-optimization in tool wear for friction-stir welding of Al-6061 + 20% Al_2O_3 MMC. Mater Sci Eng A349:156–165

Singh BR (2012) A handbook on friction stir welding: working principles, welding processes, applications, advantages and disadvantages of an advanced welding technology. LAP LAMBERT Academic Publishing GmbH & Co. KG, Saarbrucken

Sun N, Apelian D (2011) Friction stir processing of aluminum cast alloys for high performance applications. JOM 63(11):44–50

Thomas WM, Nicholas ED, Needham JC, Murch MG, Temple-Smith P, Dawes CJ (1991) Friction-stir butt welding. GB Patent No. 9125978.8, Int. Patent Appl. No. PCT/GB92/2203

Material Processing by Sliding, Grinding, Machining, Extrusion, and Wire Drawing

Contents

Abstract

In this chapter, extreme deformation phenomenon is examined in several key areas of metal working, material working science, and technology: sliding and sliding friction, grinding and machining as classical examples of subtractive manufacturing, and extrusion and wire drawing, which can involve high strains and high-strain rates, producing uniform DRX grain structures. Especially notable in sliding and grinding is the deformation zone which extends from 10 10 μm mu;m to several hundred microns, sometimes in lamellar DRX grain layers. Drawn wires, especially pearlitic steel wires, can have yield strengths in excess of 5 GPa and represent the limits of extreme deformation or severe deformation processing.

Introduction

It has been observed that many forms of severe plastic deformation or dynamic plastic deformation share a common microstructural feature: DRX. By dynamically recrystallizing, materials can respond to the deformation energy by flowing in the solid state facilitated by DRX grain boundary sliding. There are related microstructural mechanisms which augment grain refinement such as mechanical mixing of reaction products or second phases, dislocation production and arrangement, and often cracking or crack nucleation and coalescence as in localized adiabatic shear

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_52

bands illustrated schematically in ▶ Fig. 39 in chapter "Ballistic and Hypervelocity Impact and Penetration." Deformation can also extend well beyond the DRX flow zone as in the case of impact crater formation as illustrated in ▶ Fig. 16 in chapter "Ballistic and Hypervelocity Impact and Penetration," where the deformation zone extends several millimeters from the crater wall.

This chapter presents additional variations of extreme deformation. In the case of sliding wear, the deformation is limited to the near surface region and extends only ~10 μm into the bulk. This regime is narrower than an adiabatic shear band. More severe surface material displacement and deeper penetrating deformation regimes occur for grinding and machining. Even more homogeneous DRX formations can occur in swaging, extrusion, and fine wire drawing where the DRX volume can extend throughout the deformed material, particularly metals and alloys. In severe deformation cases, unwanted cracking can develop within these DRX regimes.

Sliding and Grinding

There are of course many forms of sliding of contacting surfaces, which may be the same material or two different sliding materials. The real contact area of two sliding bodies is usually much smaller than the apparent area. Surface roughness of varying degrees, which includes asperities and voids, will deform differently in response to some pressure or applied force. *Friction effects* are also present and can be affected by adhesion or cohesion as this relates to interatomic forces. Plastic deformation contributes to friction in the form of plastic work and can be expressed by

$$F\Delta x = 2E_w \sum_i V_i \tag{1}$$

where F is the friction force applied over a sliding distance, Δx, E_w is the total work expended in plastic deformation per unit volume, and V_i is the estimated volume of some surface asperity which is preferentially, plastically deformed. Two identical materials in contact each contain asperities. When these asperities are worn away or displaced, the surface is characterized by wear, and for large forces, the surface, or a much larger area of the contacting surfaces, will deform deeper into the surfaces, displacing more material and increasing the wear through cracking and fracture in the near surface regime where DRX is most dominant. Simplified expressions for the frictional force associated with plastic deformation confined to a thin surface region of thickness, t, are expressed by

$$F = 2\rho wt\overline{E}_w, \tag{2}$$

where ρ is the material density; w is the width of the deformed zone which has a thickness, t; and $\overline{E}_w$ is the mean value of the energy per unit mass within this deformed volume, wt. The plastic work can be approximated as the product of shear stress and strain (τε). The corresponding *friction coefficient,* μ, is then expressed as

$$\mu = \frac{\mathrm{wt}(\tau\varepsilon)}{\mathrm{L}}, \tag{3}$$

where L is the length of the wear track. For two dissimilar contacting (sliding) materials, Eq. 3 can be written in the form (Rigney 1988)

$$\mu = \left(\frac{\mathrm{wt}_1\tau_1\varepsilon_1}{\mathrm{L}} + \frac{\mathrm{wt}_2\tau_2\varepsilon_2}{\mathrm{L}}\right). \tag{4}$$

Wear of metals and alloys in particular involves a complex interaction of various solid-state processes, including frictional phenomena. It can involve *abrasion* and *erosion* by impacting solid or liquid particle-removing material, *fretting* involving repeated rubbing or sliding of surfaces (such as bearings), and *corrosion*, which can be combined with other mechanisms. *Galling* is another form of wear involving friction and adhesion where the material is pulled from apparent welded zones on the contacting surfaces, similar to friction-stir welding or spot welding. This occurs by DRX mixing to form a localized bond.

J.H. Dautzenberg (1980) was one of the first to identify the role of dynamic recrystallization in dry sliding wear. Because this phenomenon is near surface or surface related, sliding and wear mechanisms are influenced by surface structure, including DRX grain size. Consequently, surface grinding or polishing can produce DRX grain structures having different grain sizes and correspondingly different surface hardness. Hardness profiles can be created by wear processes which create microstructure evolution measured from the surface, including DRX evolution.

Figure 1 provides an overview of microstructure (DRX) evolution and related profiles, including hardness and equivalent strain for eutectic Ag-Cu and copper. Figure 1a can be compared with ▶ Fig. 19 in chapter "Ballistic and Hypervelocity Impact and Penetration" illustrating hardness profiles below the deformed surface, although the distances vary by orders of magnitude. Figure 1b illustrates the equivalent strain for comparison with the hardness profile in Fig. 1a, while Fig. 1c shows the grain size variance with distance from the surface. Figure 1d illustrates the wear volume variation with DRX grain size created at the surface of copper using different initial deformation techniques: dynamic plastic deformation (DPD) (Li et al. 2008), quasi-static compression, and multiple cold rolling – each method producing a cumulative strain of ~2.

The material science and engineering concerned with interacting surfaces, including friction, especially dynamic (sliding) friction, wear, and lubrication, is encompassed in the area of *tribology*. This is generally associated with very near surface phenomena as implicit in Fig. 1a–c. Deeper abrasion involves grinding processes which remove significant amounts of material from the surface, although a single depth of cut in grinding is usually only a few microns.

Grinding processes, often considered a form of machining, employ an abrasive wheel containing grains or particles of very hard material (such as Al_2O_3, carbides (SiC), or diamond, especially for fine grinding) bonded to the wheel surface in a matrix. As illustrated schematically in Fig. 2, each grain acts as an irregularly

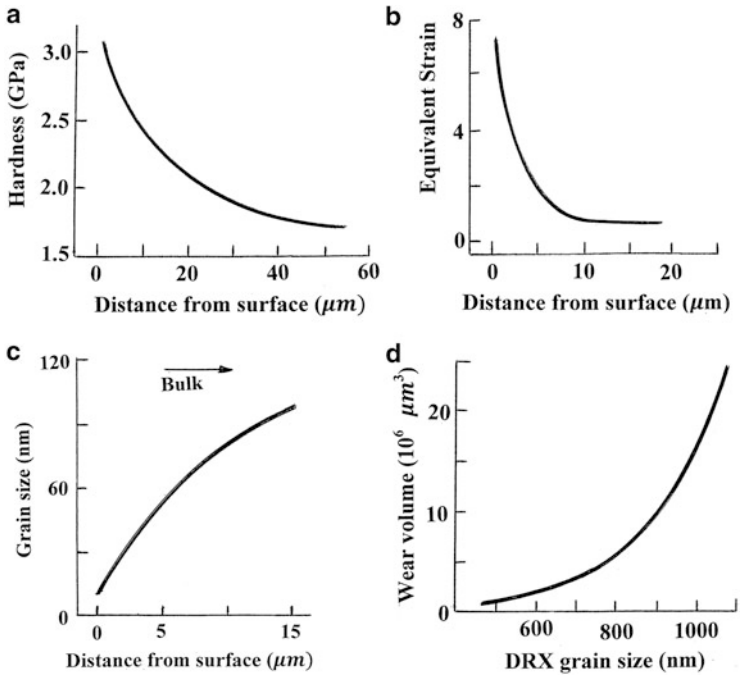


Fig. 1 (**a**) Nanoindentation hardness profile and (**b**) equivalent strain profile for eutectic Ag-Cu alloy after wear testing. Fitted to data in Cai and Bellon (213). (**c**) DRX/grain size profile for eutectic Ag-Cu alloy after wear testing. Fitted to data in Cai and Bellon (2013). (**d**) Correlation of measured wear volume with average DRX grain size for copper. Fitted to data in Yao et al. (2012). In (**c**) the bulk grain size was 114 nm

shaped cutting tool and removes an equivalent short chip as in machining (Fig. 2a). As shown in Fig. 2b, the actual grain depth of cut, t, is given by

$$t = 2\sqrt{\left(\frac{V_w}{CrV_g}\right)\sqrt{\frac{d}{D}}}, \tag{5}$$

where V_w is the workpiece velocity; C is the number of active grains on the wheel surface per unit area (usually ~1–5 mm^{-2}); $r = b'/t$ as shown in Fig. 2b for the triangular grain cross-sectional geometry; V_g is the velocity of the grinding wheel having a diameter, D; and d is the wheel depth of cut as illustrated in Fig. 2b – $d >> t$. The specific cutting energy, U, is strongly dependent on t: $U \propto 1/t$. The surface temperature in grinding, as in machining, can be significant, even leading to melting, depending upon the wheel speed and depth of cut. As in sliding wear, grinding involves DRX formation and related deformation which is projected into the material to depths which exceed wear processes, but not as far from the worked surface as in impact cratering.

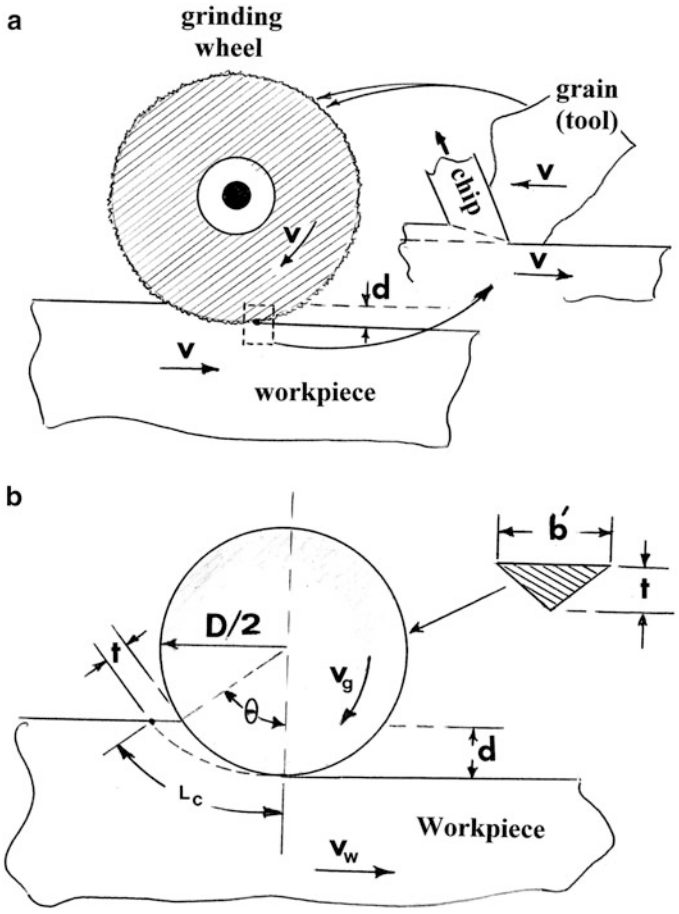


Fig. 2 Grinding process schematics. (**a**) Grinding wheel removing material from a workpiece surface to a depth d. V indicates wheel or cutting grain (tool) and workpiece velocity. The *dotted box* and *arrow* show details of local chip formation by cutting grains on wheel surface. (**b**) Details of grinding geometries, including cutting grain cross section. Specific grinding wheel velocity, V_g, and workpiece velocity, V_w, are indicated. The cut section length is L_c and the cut geometry is indicated by θ. t is the actual (individual) grain depth of cut. The workpiece depth of cut is d

Machining

In machining, a larger tool than the grinding grain (Fig. 2) removes the material near the surface of the workpiece by shearing, which forms a large chip of length L_C and thickness t_c. The actual thickness or depth of cut which is sheared by the tool is t, and the chip thickness ratio or cutting ratio is defined as $r = t/t_c$. From continuity of mass, $Vt = V_c t_c$. The sheared thickness travels along the rake face of the tool as illustrated schematically in Fig. 3. The entire chip is deformed as it meets the tool at

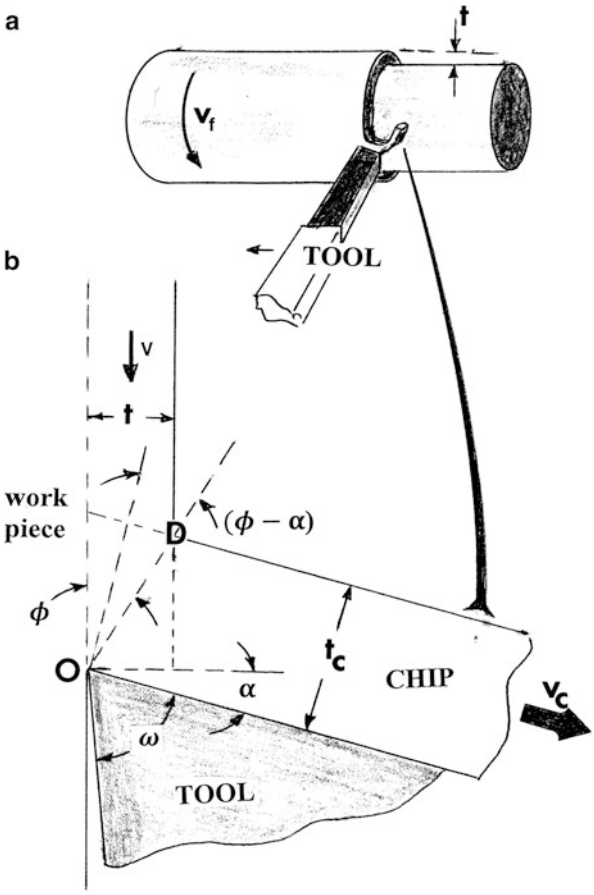


Fig. 3 Schematic views of machining mechanics. (**a**) Machining a bar section on a lathe. The feed velocity is V_f. t is the depth of cut. (**b**) Cutting geometry. The workpiece cut, t, is fed to the tool wedge at a velocity, V, forming a chip of thickness t_c which moves at velocity V_c. The shear angle, ϕ, is controlled by the cutting ratio, $r = t/t_c$. The rake angle is denoted by α. The tool wedge angle is w. Tools are specially hardened steels, carbides, etc.

a shear plane angle, ϕ. There is also a localized region of intense shear which arises in part by the friction at the rake face and referred to as secondary shear. However, this shear zone is characterized by DRX and DRX bands similar to the lamellar DRX structure-characterizing rim formation and material flow from craters and penetration channels in ballistic penetration as illustrated in ▶ Fig. 30 in chapter "Ballistic and Hypervelocity Impact and Penetration." This feature is illustrated for a 304 stainless steel chip in Fig. 4a, while Fig. 4b shows heavy deformation twins and twin faults in the primary shear zone of the chip.

The chip thickness ratio is related to the rake angle, α, and the shear angle, ϕ (▶ Fig. 3 of chapter "Fracture Modes and Mechanisms") by

$$r = t/t_c = \frac{OD \sin \phi}{OD \cos(\phi - \alpha)}. \tag{6}$$

Correspondingly, the shear strain, γ, is given by

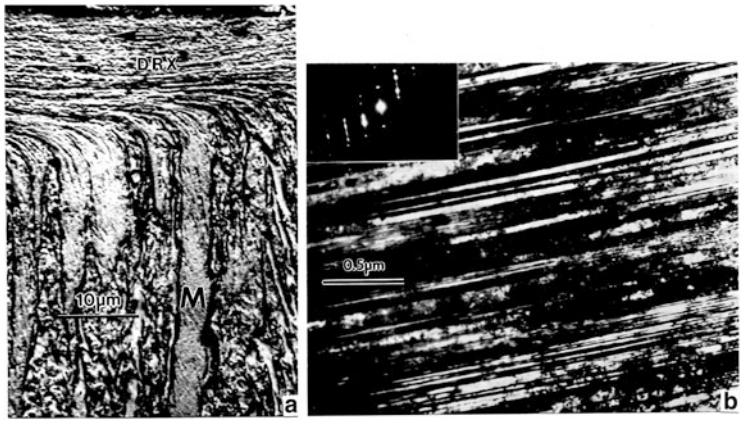


Fig. 4 Microstructure for a 304 stainless steel machined chip. (**a**) Optical microscope view of the chip section showing primary (*M*) and secondary (*DRX*) zones. (**b**) TEM image of heavy deformation twins in primary chip zone at M in (**a**). The selected area electron diffraction pattern *insert* shows twin reflections and associated diffraction streaks perpendicular to the twin plane. The grain surface orientation is (110)

$$\gamma = \cos\alpha/\sin\phi\cos(\phi - \alpha). \tag{7}$$

The shear strain rate in machining is

$$\dot{\gamma} = \frac{d\gamma}{dt} = \frac{V_s}{(Y_s)_{max}}, \tag{8}$$

where $(Y_s)_{max}$ is the estimated maximum value of the shear zone thickness (often estimated to be ~25 mm) and the shear velocity, V_s, is given by

$$V_s = V\cos\alpha/\cos(\phi - \alpha),$$

where V is the velocity of the tool relative to the workpiece in Fig. 2b. Common strain rates are $\sim 10^2\ s^{-1}$, and efficient machining occurs when $\phi \cong 45°$. The average shear stress τ is given by the shear force, F_s (Fig. 2b), divided by the shear plane area, A_s, where $A_s = bt/\sin\phi$ and b is the width:

$$\tau = F_s/A_s = F_s\sin\phi/bt. \tag{9}$$

It is the shear stress which mainly affects the energy requirement in machining.

In contrast to a cutting tool as illustrated in Figs. 2 and 3, material removal can be accomplished by *milling,* which uses a rotary cutting tool which is stationary relative to the advancing workpiece but which can have up to 5° of freedom or axes. The blades of the cutter are organized in contrast to the random grinding tool particles illustrated on the grinding wheel in Fig. 2. The cutter shaves multiple chips from the surface with each revolution. Modern milling centers are *computer*

numerical control (CNC) milling machines which, as noted above, have two more axes in addition to the normal x, y, z axes (Groover 2007). Surface region deformation is similar to grinding and machining, but the depth of cut is more like machining and the propagation of microstructure commensurate with machining relative to the prior cut surface.

Metal removal by grinding, machining, or milling is subtractive manufacturing in contrast to additive manufacturing treated in chapter "▶ 3D and Multidimensional Materials Science." Correspondingly, the development and evolution of microstructure is considerably different. Additive manufacturing microstructures exhibit a diversity dependent upon heating-cooling cycles, layer thickness, etc. in contrast to subtractive manufacturing microstructures which are variances of DRX grain or lamellar structures which evolve into varying densities and configurations of dislocations or related deformation microstructures such as twinning as illustrated in Fig. 3b.

Extrusion and Wire Drawing

In extrusion and wire drawing, the material is shaped by a die which deforms the material as it is forced through the die. The strain involved in the deformation is related to the variation in cross-sectional area or dimension on passing through the die as illustrated schematically in Fig. 5a. The corresponding strain rate, with reference to Fig. 5a, is given approximately by

$$\dot{\varepsilon} \cong \frac{\ln(d_o/d')v}{L}, \tag{10}$$

where d_o is the initial diameter, d' is the drawn or extruded diameter, L is the approximate extent of the plastic deformation zone relative to the die, and v is the extrusion or drawing velocity.

Because there may be subtle differences in the work hardening or deformation characteristics as a material is extruded at high-strain or strain rate conditions, there may be variances in the uniformity of solid-state flow, especially near the central axis of drawing. In addition, velocity gradients may occur, and these would also create flow variances. Surface (die) friction can also influence these phenomena. However, when such phenomena occur, internal void-related defects can occur as illustrated in Fig. 5b, c. The common void form is chevron shaped which point in the direction of metal flow. These are also referred to as central-burst phenomenon. Severely cold-worked or low ductility materials exaggerate this effect. This defect can be reduced or eliminated by increasing the reduction (or strain) or decreasing the die half-angle. Correspondingly, the precursor material can be annealed.

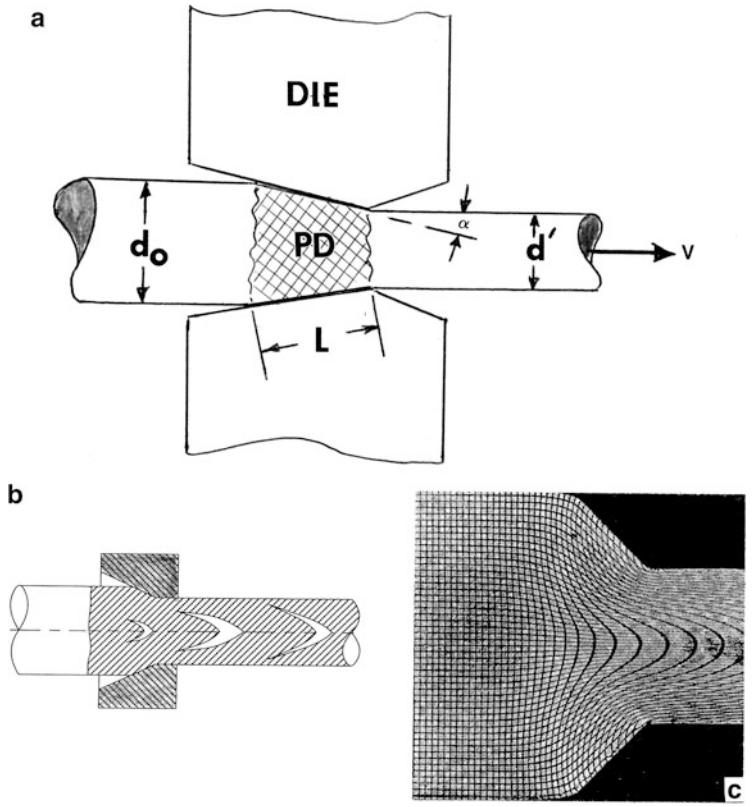


Fig. 5 Extrusion or drawing die schematic (**a**) and chevron-like defect formation in drawing (**b**). In (**a**) the die angle is denoted by α. PD indicates the plastic deformation zone having a length L along the die face. The drawing velocity is V. (**b**) Shows a schematic for chevron central-burst void defect formation. (**c**) Distorted grid pattern in axisymmetric extrusion of a Pb-Zn soft alloy using a die lubricant

For very small wire drawing where the reduction is high and the corresponding strain is large, inclusions such as oxides in the material can also create void-like features at these inclusions, and the drawn microstructure can contain strings of voids and inclusions oriented along the rod or wire axis, even when the deformation is uniform as illustrated schematically in Fig. 6a. Figure 6b, c illustrate the directional void features and related microstructural texturing in drawn copper rod and wire. Figure 7a, b illustrate the microstructures corresponding to the copper rod in Fig. 6b which was continuously cast and drawn to 7.9 mm (5/16 in.) in a series of dies. Figure 7c, d show copper wire drawn to 46 gauge (~40 μm in diameter) where the entire cross section is

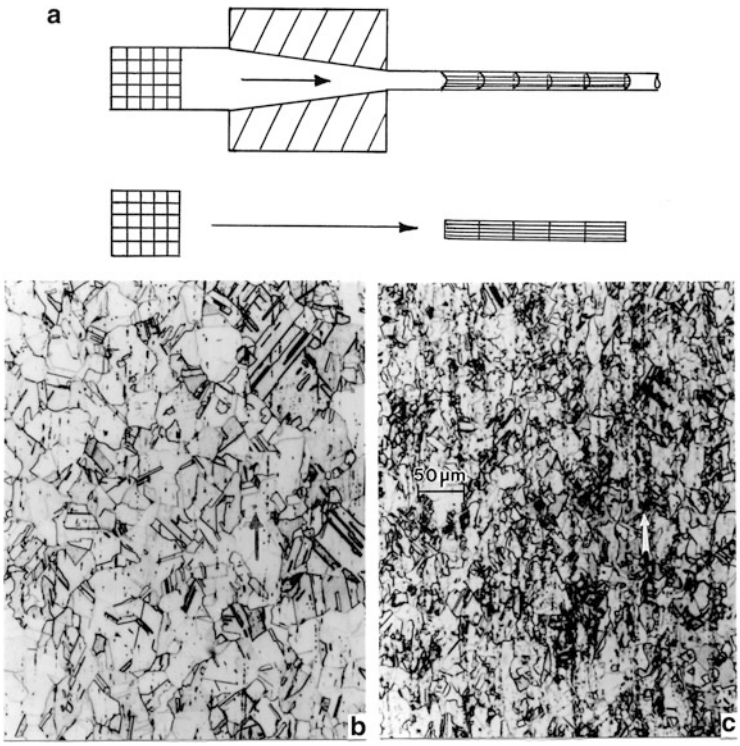


Fig. 6 Grid distortion schematic for wire drawing (**a**) and residual, drawn copper rod (**b**), and wire (**c**). In **b**, the rod diameter was 7.9 mm, while in **c** it was 2 mm. The drawing direction and wire axis are shown by *arrows* in **b** and **c**. Magnification in **b** is shown in **c**

composed of elongated (textured) DRX grains ~90 nm in diameter. The selected area electron diffraction pattern insert in Fig. 7d is indicative of these nano-grain features.

During cold wire drawing of very fine wires as in Fig. 7c, d, the strain rate can reach the range of 10^3–10^4 s^{-1}, and corresponding true strains can exceed 3–5, in some cases even as high as 8 or 9. The drawing of steel wires in particular under such extreme conditions, and with requisite microstructures, especially carbides, can produce wire products with extremely high strengths. For example, pearlitic steel tire cord can have tensile strengths near 5 GPa, and pearlitic steel drawn and braided to form bridge cables has a tensile strength near 2.5 GPa (Raabe et al. 2010).

It should of course be recognized from Fig. 7 in particular that extreme deformation, or severe plastic deformation (SPD) discussed in chapter "► Synthesis and Processing of Nanomaterials," is also an effective route for producing bulk

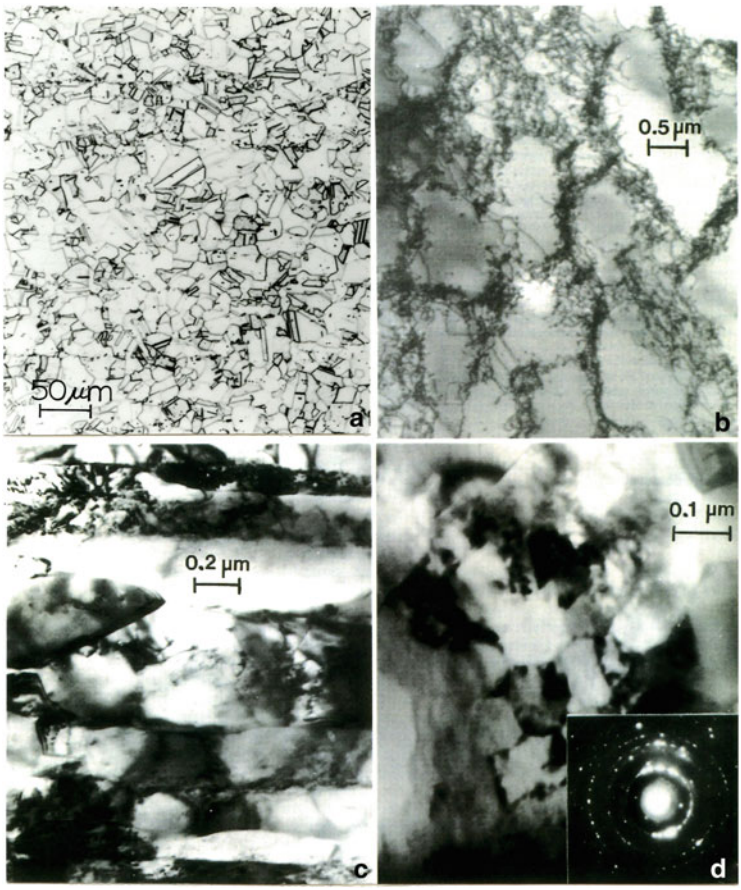


Fig. 7 (**a**) Light microscope view along a longitudinal half-section for a commercial Cu rod, 7.9 mm in diameter as in Fig. 6 (**b**), with a 15 μm grain size. (**b**) TEM image showing dislocation cells corresponding to **a**. (**c**) TEM image of a longitudinal wire section (~40 μm in diameter) showing elongated nanograins. (**d**) TEM image and selected area electron diffraction pattern insert for corresponding, transverse view in **c**; looking into the wire axis

nanocrystalline materials. These can include equal channel angular extrusion and friction-stir processing as illustrated schematically in ▶ Figs. 11 and ▶ 12 of chapter "Synthesis and Processing of Nanomaterials," respectively.

References

Cai W, Bellon P (2013) Subsurface microstructure evolution and deformation mechanism of Ag-Cu eutectic alloy after dry sliding wear. Wear 303:602–610

Dautzenberg JH (1980) The role of dynamic recrystallization in dry sliding wear. Wear 60:401–411

Groover MP (2007) Theory of metal machining. In: Groover MP (ed) Fundamentals of modern manufacturing, 3rd edn. Wiley, New York, pp 491–504

Li YS, Tao NR, Lu K (2008) Microstructure evolution and nanostructure formation in copper during dynamic deformation at cryogenic temperature. Acta Mater 56:230–241

Raabe D, Choi P-P, Li Y, Kostka A, Saurage X, Lecoutarier F, Hono K, Kirchheim R, Pippan R, Embury D (2010) Metallic composites processed via extreme deformation: toward the limits of strength in bulk materials. MRS Bull 35:982–991

Rigney DA (1988) Sliding wear of metals. Ann Rev Mater Sci 18:141–163

Yao B, Han Z, Lu K (2012) Correlation between wear resistance and subsurface recrystallization structure in copper. Wear 294–295:438–445

Part XIV

Materials Degradation and Failure

Fracture Modes and Mechanisms

Contents

Abstract

Breaking bonds of any type is the fundamental precursor to crack nucleation. Catastrophic brittle or cleavage fracture produces essentially only cleavage steps, while variations in ductile fracture produce different sizes of cup-cone (ductile-dimple) fracture surface structures. Steels in particular are prone to low-temperature ductile – brittle transitions while variations or cycles in stress produce fatigue structures and behavior. Fracture induced by creep, fatigue, creep-fatigue, impact, as well as ductile/brittle fracture modes are illustrated. Crack nucleation, coalescence, and growth are described in detail along with crack initiation sites corresponding to the various fracture modes. Complex interactions of fracture modes in various stress, temperature, and chemical environments are illustrated.

Introduction: Fracture Fundamentals

Fracture represents the failure end point: the continuous breaking of atomic bonds leading to (catastrophic) separation of a solid material. Nucleation, growth, and coalescence of cracks or void precursors represent prior stages of degradation or pre-failure. These processes are different in different materials: crystalline or noncrystalline metals and alloys, ceramics, polymers, semiconductors, etc. or materials exhibiting different bonding characteristics, especially characterized by different atomic bonding, atomic arrangements, crystal structures, and other structural phenomena; including defects in crystal structures. These include amorphous and crystalline

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_53

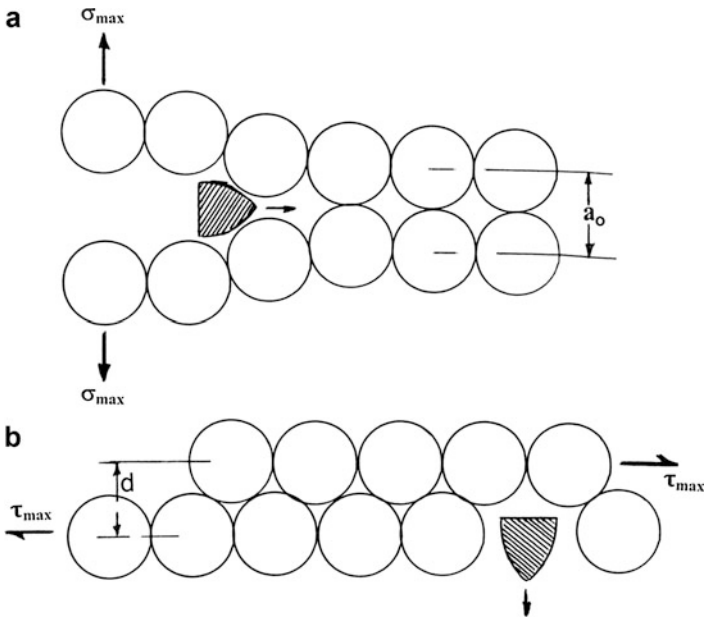


Fig. 1 Tensile (**a**) and shear (**b**) mechanisms for atomic separation creating incipient crack nuclei

metals and alloys, ceramics, polymers, and composites of these as the matrix. Fracture in these systems depends upon modes of deformation at temperature: tension (uniaxial and biaxial), uniaxial compression, high-strain rate uniaxial tension, impact, fatigue fracture, and creep rupture. These can span a strain-rate range from $<10^{-6}\ s^{-1}$ to $>10^{8}\ s^{-1}$, or ~15 orders of magnitude.

In the simplest approach, fracture of crystalline solids may be represented by either the separation of atoms by bond rupture or by atomic shear displacements causing separation. In the latter it is apparent that dislocation shear displacements are precursors to rupture by shear interruption. Correspondingly, there are two different forces affecting these two separation mechanisms (Fig. 1):

$$1:\ \text{Breaking stress};\ \sigma_{max} = \sqrt{E\gamma/a_o} \tag{1}$$

$$2:\ \text{Theoretical shear stress};\ \tau_{max} = Gb'/2\pi d \tag{2}$$

where E is Young's modulus, γ is the surface free energy, a_o is the equilibrium spacing of the atomic planes, G is the shear modulus, b' is a lattice parameter equal to $a/\sqrt{6}$ for fcc metals, and d is the interplanar spacing parallel to the applied shear stress, τ. For common fcc metals (Cu, Ni, Al, etc.), $\sigma_{max} \simeq E/15$ and $\tau_{max} \simeq G/15$, where $E = 2(1-\upsilon)\,G$ and υ is Poisson's ratio (~0.3 for the metals indicated). Substituting values for E and G shows that shear requires a smaller stress for separation. However, by lowering γ, the situation can reverse. To some extent, Eqs. 1 and 2 represent brittle versus ductile crack (fracture) initiation at the atomic level.

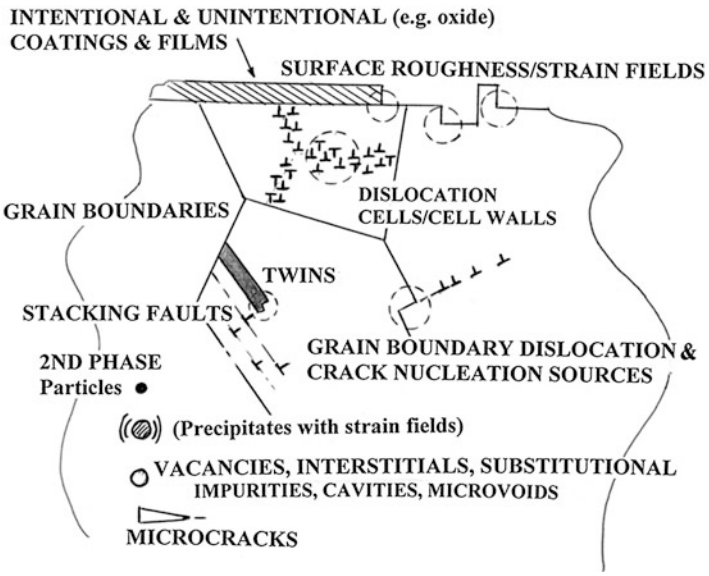


Fig. 2 Microstructural regimes contributing to ductile fracture in metals and alloys (Adapted from Wilsdorf (1983))

In contrast to brittle fracture, ductile fracture is considerably more complex because there are numerous microstructural features which can create shear instabilities from which microvoids and microcracks can nucleate. These are illustrated schematically in Fig. 2 which is adapted from the classic treatment of ductile fracture of crystalline or polycrystalline metals by Heinz G.F. Wilsdorf (1983). Figure 2 lists essentially all microstructural phenomena in crystalline materials since under the conditions illustrated in Fig. 1 all can initiate or nucleate cracks or crack precursors: voids. Corresponding to these microstructural features, Fig. 3 shows more fundamentally how microcracks can be initiated at dislocation instabilities arising at intense dislocation pileups, which can alternatively create a ledge at an interface. Extremely high-density dislocation structures such as dislocation wall structures in dislocation cells (Fig. 2) formed at very high stresses (or strains), and intergranular and transgranular microcrack initiation at interfacial strain instabilities such as ledges or surface steps can also form as illustrated ideally in Fig. 3. Such shear instabilities can occur in dynamic recrystallization regimes in tensile necking near the fracture stress or in more localized shear instabilities associated with adiabatic shear bands reproduced in Fig. 4 showing crack nucleation which can occur in the DRX grain boundaries or in the grain interiors according to Fig. 3. Figure 4 also illustrates the growth and coalescence of microvoids to form a propagating microcrack. These features can occur in highly dislocated zones such as dense dislocation cell walls or along heavily deformed grain or phase boundaries.

In ductile fracture crack initiation, there are three simple crack-opening modes as illustrated in Fig. 5. These are variances on the two separations, and crack

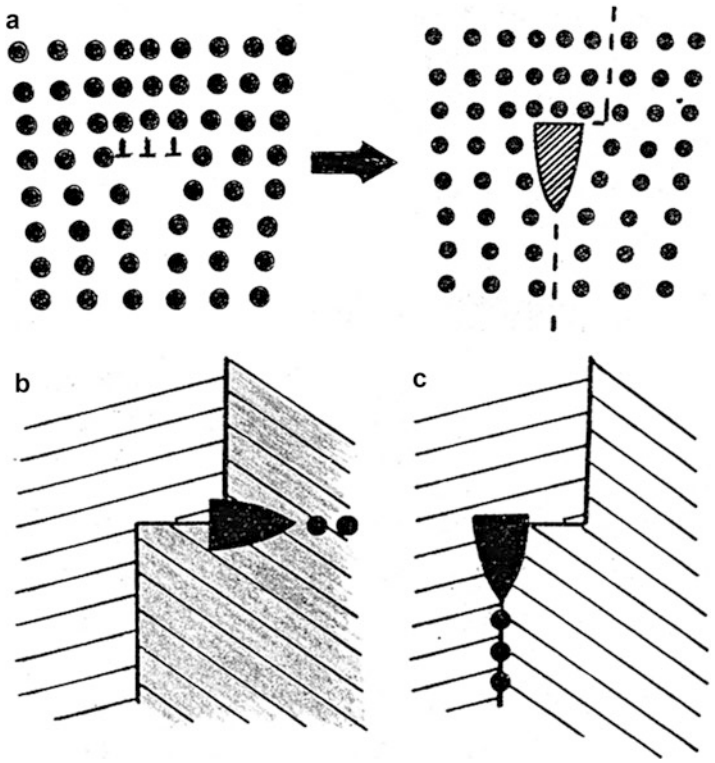


Fig. 3 Microcrack nucleus formation. (**a**) Edge dislocation coalescence creating microledge. (**b**) Transgranular-directed microcrack. Shading separates two crystals with interfacial ledge. (**c**) Intergranular- or interfacial-directed microcrack in contrast to (**b**) (Adapted from Murr 1975)

nucleation mechanisms shown in Fig. 1, and to some extent emulate the topological (Volterra) defects illustrated in ▶ Fig. 1 of chapter "Line Defects: Dislocations in Crystalline Materials". The behavior of a crack is determined empirically by a stress intensity factor, K, or stress field which upon reaching a critical value, K_c, crack propagation is imminent. Generally

$$K = \sigma_n \sqrt{2\pi r}, \tag{3}$$

where σ_n is the nominal stress and r is the distance from the crack tip. For the condition of plane stress

$$K = K_{\bar{\sigma}} = \sqrt{FE}, \tag{4}$$

where F is the strain energy release rate with crack extension (or the crack extension force) and E is the Young's modulus. For plane strain,

$$K = K_{\bar{\epsilon}} = FE/1 - \upsilon^2, \tag{5}$$

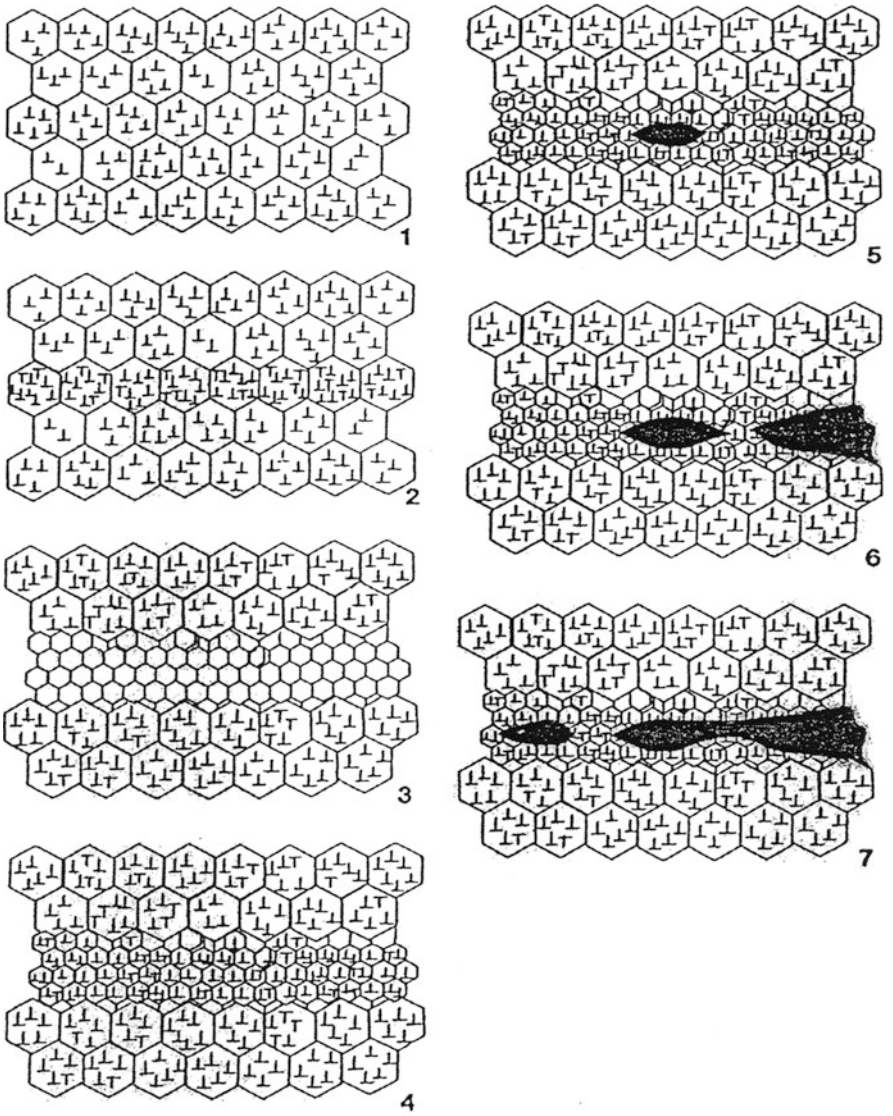


Fig. 4 Formation, coalescence, and growth of microcracks in DRX zone formed in adiabatic shear instability regime. DRX zone formed in 3 is deformed in 4 creating crack (void) nucleus in 5, 6 and 7 show void coalescence and crack growth

where

$$F = \pi\sigma^2 c/E, \tag{6}$$

σ in Eq. 6 is the applied stress and c is half the crack length. The corresponding crack tip–opening displacement, δ, which is measured at the elastic–plastic interface of the crack, is defined by

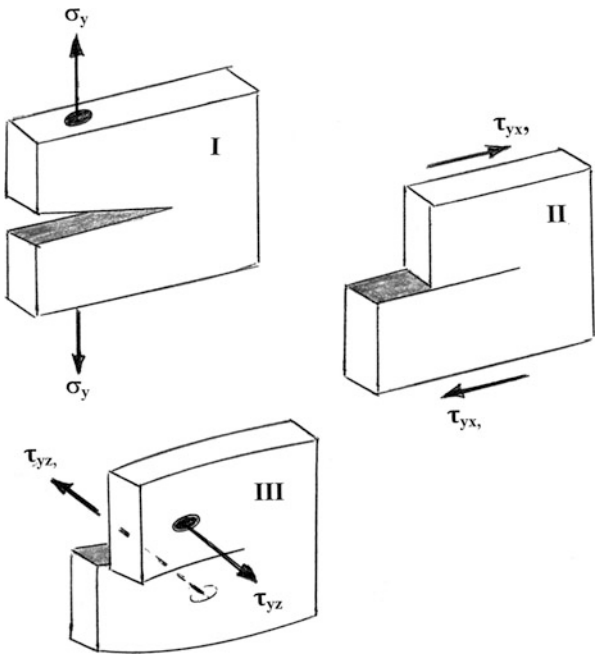


Fig. 5 Simple fracture (crack-opening) modes: I tensile; II shear; III tearing or anti-plane (shear) mode

$$\delta = K^2/\sigma_y E, \tag{7}$$

where σ_y is the yield stress. As the ductile crack propagates, a plastic (deformation) zone must form in front of the crack tip with an estimated radius, r_p, where

$$r_p = K^2/2\pi\sigma_y{}^2 = FE/2\pi\sigma_y{}^2. \tag{8}$$

It is perhaps obvious that crack nucleation in ductile metals and alloys will occur at a region of shear instability or stress (strain) concentration as illustrated in Figs. 2, 3, and 4. Consequently highly plastic (deformed) zones will nucleate a crack, and crack tips themselves will create plastic zones at some distance from the advancing crack tip by dislocation emission through a kind of strain concentration. The opening crack also creates highly strained regions next to the crack surface (or edges), including DRX zones. These features have been observed in thin films strained in tension in transmission electron microscopes (Wilsdorf 1983; Murr 1982). Figures 6 and 7 illustrate high-voltage TEM sequences showing crack nucleation, coalescence, and growth (propagation) in thin films of 304 stainless steel (having an average grain size of ~10 μm), strained in tension (in situ). These films were ~ one-tenth-grain thick and not electron transparent in the sections viewed as shadowgraphic (mass-thickness contrast) images in Figs. 6 and 7 (Murr 1991). Consequently, the images emulate the schematic features illustrated two-dimensionally in Fig. 4, especially Fig. 7f. Figures 8 and 9 compliment the dynamic crack images in Figs. 6 and 7 by providing

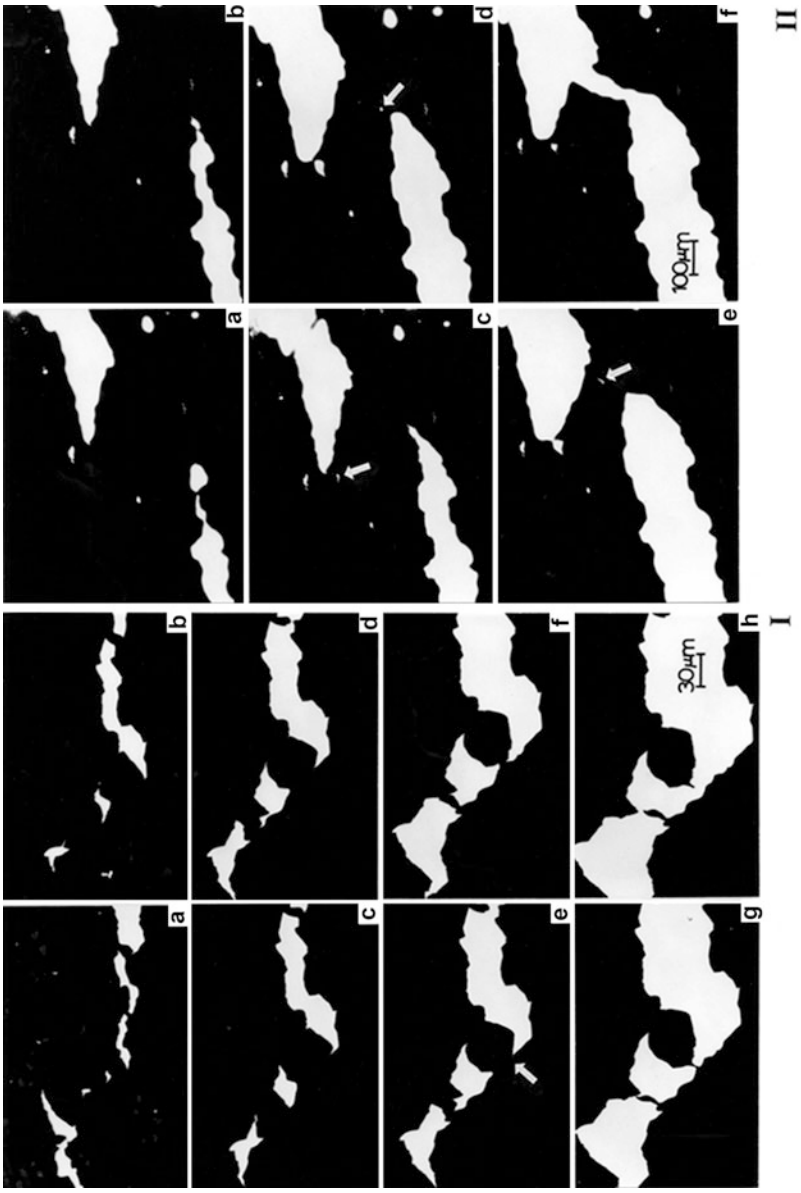


Fig. 6 Observations of ductile fracture progression in thin 304 stainless steel specimen stressed in a high-voltage (1 MV) transmission electron microscope. I: Note crack coalescence especially from (**e**) to (**h**). II: Note coalescence and crack growth (*arrows*) in (**c**) to (**f**)

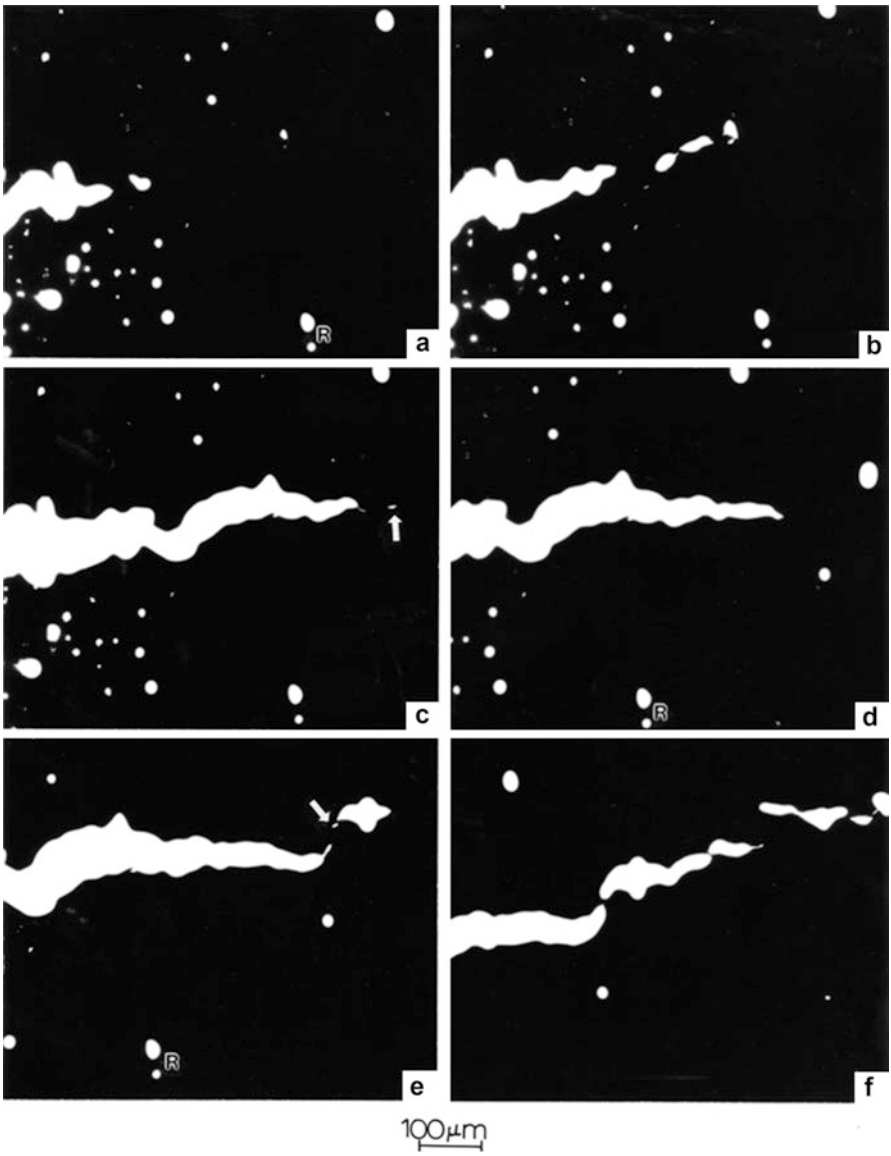


Fig. 7 Enlarged view of crack coalescence and growth in 304 stainless steel deformed (in situ) in a high-voltage (1 MV) TEM. Note precursor void at arrow in (**c**). R in (**d**) and (**e**) shows reference area. Note precursor void at arrow in (**e**) and coalescence and growth in (**f**)

some crack-related microstructural features in thinner (electron transparent) areas of the in situ-stressed thin 304 stainless steel films in Figs. 6 and 7. It is enlightening to view Figs. 6, 7, 8, and 9 somewhat systematically in order to form a mental concept for ductile crack nucleation, coalescence, and propagation.

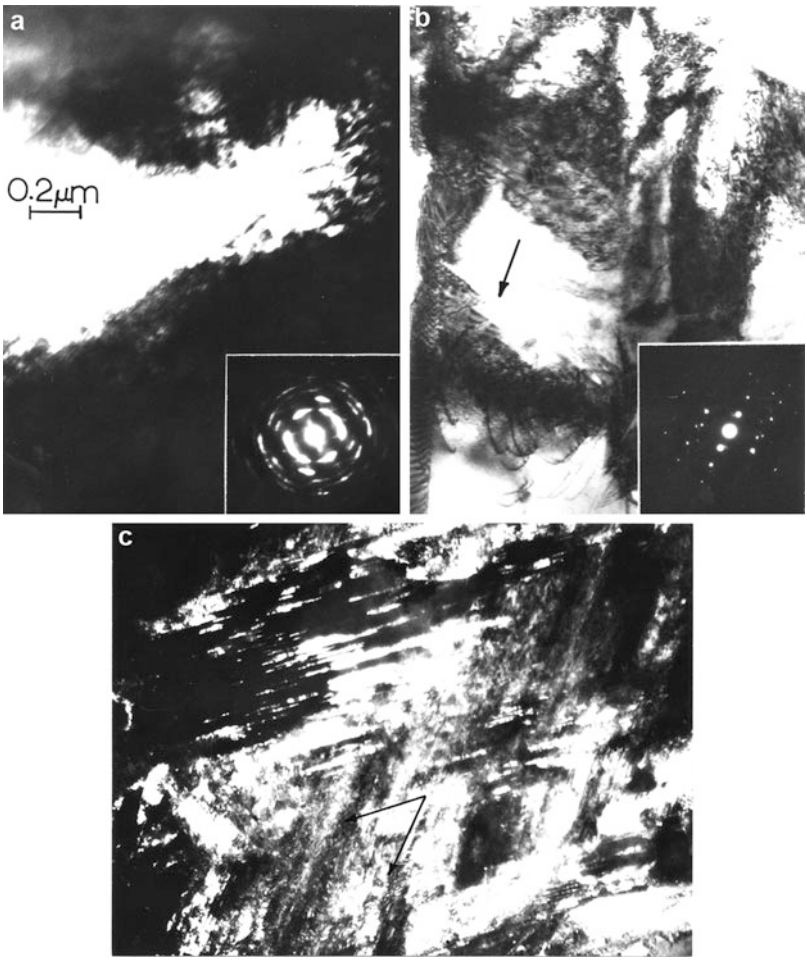


Fig. 8 TEM images showing crack tip and crack edge thinning and microstructure development corresponding to crack growth progression shown in Figs. 6 and 7. (**a**) Crack tip region. Selected-area electron diffraction pattern insert shows DRX grains and deformation twins. (**b**) Crack nucleation (*arrow*) in heavily deformed region. (**c**) Dark-field image showing deformation twins and transformed α′-phase in 304 stainless steel crack edge region. *Arrows* show <112> directions in (110) fcc surface orientation

Figure 8a shows a thinned region of highly deformed material surrounding a crack tip which includes DRX grains, twins, and transformed α′ (bct) phase indicated in the selected-area electron diffraction pattern insert. Figure 8b shows a crack nucleating in a heavily dislocated zone. Figure 8c shows a highly deformed region near a crack edge containing α′-phase plates and deformation twins coincident with fcc {111} planes (arrows). Figure 9 shows a schematic view of a propagating crack incorporating the microstructures shown in Fig. 8; which can include intermixed DRX grains. The involvement of DRX grains, especially the

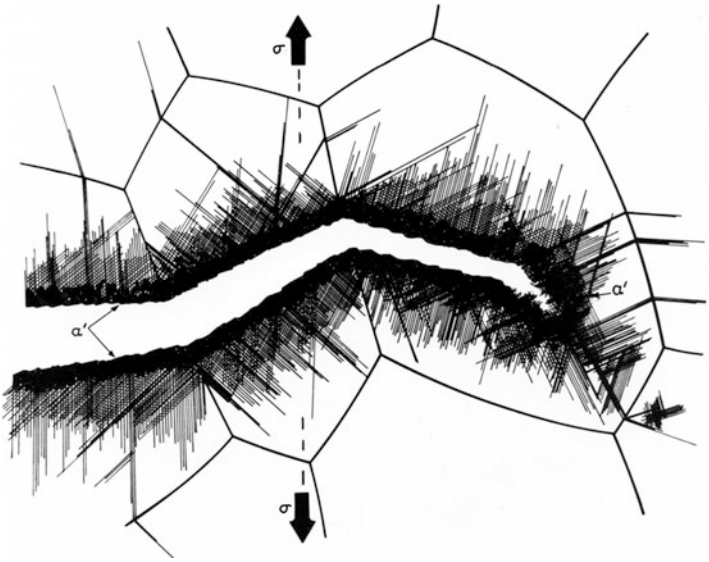


Fig. 9 Crack propagation and edge microstructure composed of {111}-related deformation twin-faults creating α′-phase (bct). The applied stress is indicated by σ

formation of nanosize DRX grains in heavily deformed zones in ductile metals and alloys, poses special circumstances because the associated grain boundaries or grain boundary regions are wider and more disordered than high-angle (or high energy) grain boundaries as illustrated in ▶ Fig. 8 of chapter "Mechanical and Other Properties of Nanocrystalline Materials" and recently discussed in the context of the generalized Hall–Petch relationship by Armstrong (2012). Figure 10 shows a ball-bearing model comparing nanograins and associated grain boundaries with the evolution of larger grains and correspondingly more organized and thinner grain boundary regimes. It can be noted that vacancies in the nano regime (Fig. 10 A) occur in a fivefold atom array (f) in contrast to a close-packed lattice vacancy (V) with sixfold atom arrays. In 3-dimensional solids, this would correspond to an icosahedral vacancy in contrast to a cuboctahedral or truncated cuboctahedron. Note that these vacancy arrays tend to be somewhat quantitative on comparing Fig. 10 A to D, where Fig. 10 A contains more "f" type vacancies than Fig. 10 D, for example.

While crack nucleation, coalescence, and growth in ductile metals and alloys can be influenced by grain size and grain boundary structures as implicit in ▶ Figs. 10 and 47 of chapter Mechanical and Other Properties of Nanocrystalline Materials, DRX nanograin structures can also allow for grain boundary sliding which can produce void regions and intrinsic microcrack tips shown schematically in Fig. 3d, e. At lower strain rates characteristic of creep, or in low-cycle fatigue, dislocation interactions producing atomic vacancies can contribute to grain boundary crack nucleation because of their preferential migration (or diffusion) along the grain boundaries, which is also enhanced along nanograin boundaries implicit in Fig. 10, where the vacancy concentration is larger. Temperature also contributes significantly as the

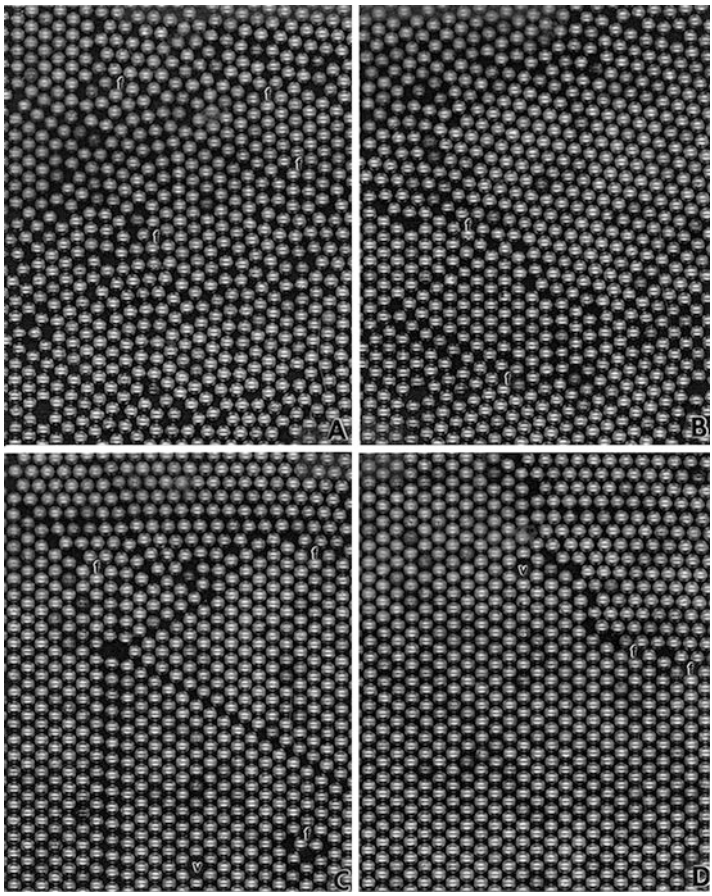


Fig. 10 Ball-bearing models for close-packed, 2-dimensional grain structures. Nanograin and grain boundary structure in A increases from B to D. Close-packed vacancies are denoted "v." Fivefold symmetric vacancies are denoted "f"

strain rate is reduced in the creep regime ($\dot{\varepsilon} < 10^{-3}s^{-1}$). Ashby (1983) has constructed fracture mechanism maps similar to the deformation mechanism maps shown in ▶ Fig. 5 of chapter "Examples of Directional Crystal Structures: Gas-Turbine Component Applications in Superalloys". Figure 11 illustrates this concept for nickel deformed by shear (▶ Fig. 11a of chapter "Electrovalent Crystal Structures and Chemistry") and fractured in tension (Fig. 11b) which shows fracture-type regions based on corresponding fracture surface observations. Fatigue fracture is not represented in this graphical format, but fatigue fracture surface features exhibit unique striations to be discussed later.

Figure 11 compares the generalized deformation map (Fig. 11a) with the corresponding fracture mechanism map (Fig. 11b) for fcc nickel. It can be observed that creep deformation and corresponding creep fracture (transgranular and especially

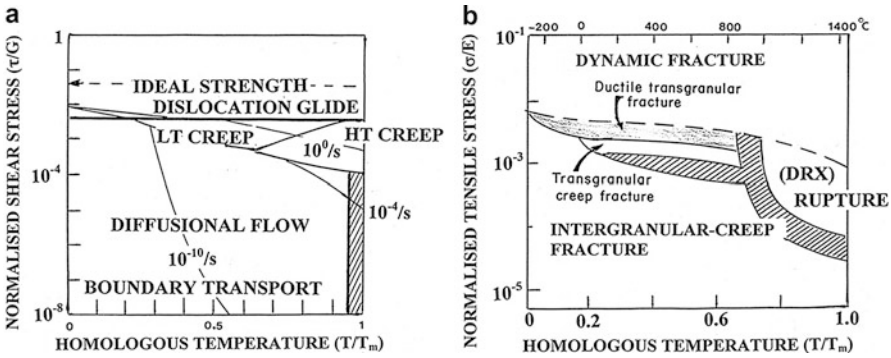


Fig. 11 Deformation mechanism (**a**) and fracture mechanism (**b**) maps for 10 μm grain size Ni. The shaded zone to right in (**a**) represents bulk (lattice) transport. Creep rates are noted between 10°/s and 10^{-10}/s. T_m is the melting temperature (1,420 °C). In (**a**), LT and HT represent low-temperature and high-temperature, respectively. Strain-rate contours and field boundaries will shift with grain size. Shading in (**b**) shows mixed fracture mode (Adapted from Ashby (1983))

intergranular) are particularly prominent at lower homologous temperatures and normalized stresses. Note that in Fig. 11b, the shading represents a mixed mode of fracture: ductile transgranular-to-transgranular creep fracture-to-intergranular creep fracture at a homologous temperature $T/T_m = 0.5$, where T_m is the melting temperature for Ni (~1,420 °C). Above about $\sigma/E = 10^{-2}$ in Fig. 11b, the fracture is characterized by dynamic fracture.

Although creep mechanisms have been discussed earlier in chapter "▶ Examples of Directional Crystal Structures: Gas-Turbine Component Applications in Superalloys" in connection with deformation mechanism map construction (▶ Fig. 5 of chapter "Examples of Directional Crystal Structures: Gas-Turbine Component Applications in Superalloys"), it would seem useful to briefly review generalized concepts and mechanisms as these relate to creep fracture – which as illustrated in Fig. 11b is both stress and temperature dependent. Dislocated glide, particularly thermally activated glide, climb, and climb-assisted glide are perhaps implicit in Fig. 3a, while bulk and grain boundary diffusion also play a major role at low creep rates or creep strain rates. As noted in chapter "▶ Examples of Directional Crystal Structures: Gas-Turbine Component Applications in Superalloys," bulk (volume) diffusion D_v is characteristic of Nabarro-Herring creep, while grain boundary diffusion, D_{gb}, is characteristic of Coble creep. Dislocations can also contribute to diffusion when the dislocation cores act as pipes; dislocation pipe diffusion, D_p. Diffusional phenomena are both temperature and structure dependent:

$$D_v < D_p < D_{gb}.$$

Since grain boundaries are sinks for vacancies, dislocations in the grain matrix can pump vacancies to the grain boundaries, or if the dislocations form dense cell wall structures, these structures can aggregate vacancies and form voids as crack

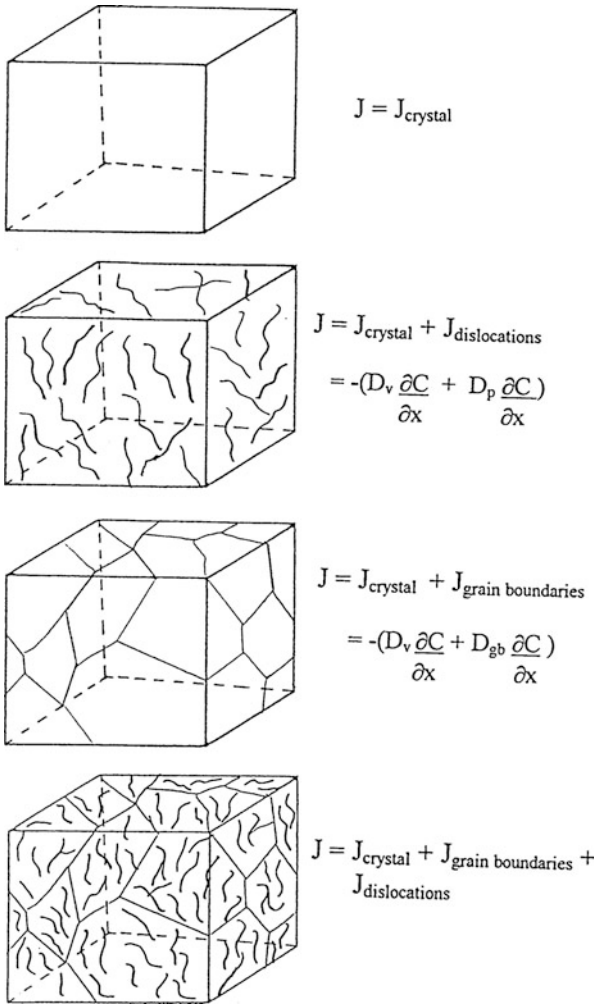


Fig. 12 One-dimensional (x-axis) flow (J) of diffusing species (including vacancies) in crystal volumes. From *top*: single-crystal; dislocations in crystal volume; polycrystal volume containing grain boundaries; dislocated, polycrystal volume

precursors. Ideally, such transport phenomena can apply not only to vacancies but also to other diffusional species, such as the transport of Cr and C to grain boundaries in stainless steels to form carbides in the grain boundaries, leading to sensitization of the boundaries by Cr depletion which eliminates the passivation at the grain boundaries and makes them sensitive to selective corrosion.

These diffusional or transport mechanisms can be illustrated generally as shown in Fig. 12 where the progressive contribution of specific diffusional parameters (diffusivity) is represented. Each type of diffusivity can be described by a simple one-dimensional version of Fick's first law of diffusion:

$$J_i = -D_i \frac{\partial C}{\partial x} \tag{9}$$

and

$$D_i = D_{o(i)}\, e^{-Qi/RT}, \tag{10}$$

where (i) represents the specific diffusion type: bulk or volume diffusion in the lattice (D_v), grain boundary diffusion, D_{gb}, and dislocation pipe diffusion, D_p. In Eq. 9, C is the concentration of the diffusing species, vacancies in the case of creep. $D_{o(i)}$ in Eq. 10 is the jump frequency, while Q_i is the activation energy corresponding to the transport mechanism, R is the gas constant, and T is the absolute temperature. A generalized creep equation considering Fig. 12 might take the form

$$\frac{d\varepsilon}{dt} = \dot{\varepsilon} = \frac{A\sigma V_m}{RTD^2}\left(D_v + \frac{\pi \Delta t}{D} D_{gb} + \rho D_p\right) \tag{11}$$

where A is a constant σ is the applied stress, V_m is the molar volume, D is the grain size, Δt is the grain boundary width, and ρ is the dislocation density. Alternatively the creep rate has been expressed by

$$\dot{\varepsilon} = B\left(\frac{\sigma^m}{D^n}\right) e^{-Q/kT}, \tag{12}$$

where B is a material parameter, σ is the applied stress, D is the grain size, k is Boltzman's constant, T is the absolute temperature, and m and n are exponents which depend upon the creep mechanisms dominating mechanism. At high stress, σ, and considering dislocation creep (or creep controlled by dislocation glide) (Fig. 11a), Q in Eq. 12 represents the activation energy for self-diffusion and m ranges from 4 to 6 while $n = 0$, no grain size dependence. Correspondingly, for Nabarro-Herring creep, $m = 1$; $n = 2$. For Coble creep, Q represents the activation energy from dominant grain boundary diffusion and $m = 1$; $n = 3$ in Eq. (12). These represent different weightings of diffusivities, etc. in Eq. 11 as well.

As illustrated in Fig. 10 and discussed by Armstrong (2012), Δt (Eq. 11) is larger for nanograin materials, while the effective interfacial area increases dramatically, making D_{gb} dominant for creep and creep fracture in nanomaterials. This phenomenon is illustrated in several similar perspective views shown in Fig. 13 which also compares deformed and undeformed large-grain and nanograin structures. Figure 13a shows large-grain boundaries with no significant grain boundary width, while Fig. 13b is indicative of a grain boundary regime for both the large-grain and nanograin perspectives.

It is perhaps expedient if not simply a matter of completeness to further extend this fracture introduction to include fatigue and fatigue failure (and fracture). Fatigue in its broadest sense involves the mechanical degradation and eventual fracture of a material as a consequence of *cyclic loading*, which can include thermal cycling or cyclic thermal stresses. Typically failure under cyclic loading occurs at stress levels which are often much lower than failure stresses for monotonic loading. Figure 14a illustrates the basic fatigue parameters corresponding to cyclic

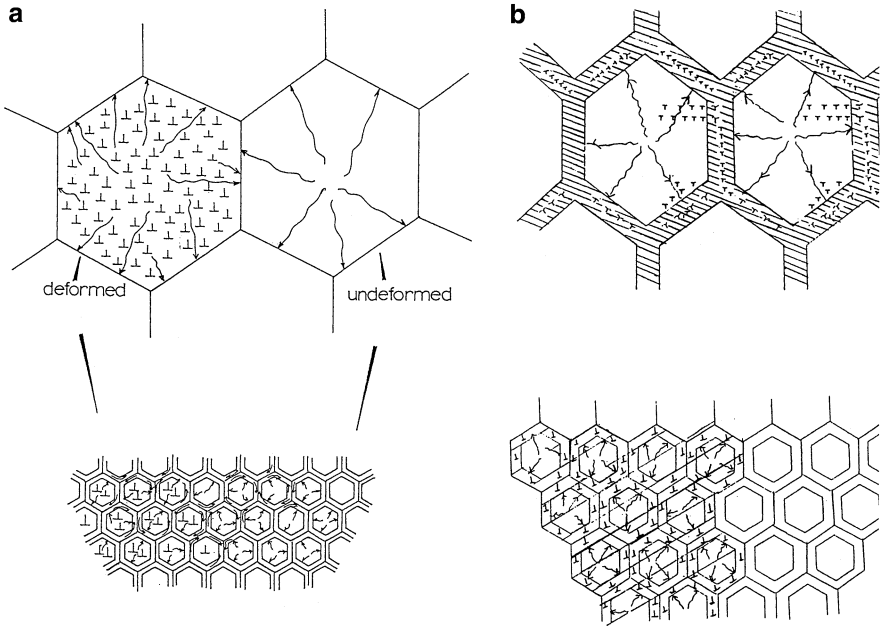


Fig. 13 Schematic views showing conventional grain structures in contrast to nanograin structures. (**a**) Deformed versus undeformed grain structures. Note grain boundary widths relative to grain (crystal matrix) size. (**b**) Grain and grain boundary dislocation structure in conventional and nanograin regimes

loading, where σ_{max} and σ_{min} are the maximum and minimum stress levels, respectively. Figure 14b illustrates the typical representation of fatigue life of S (stress) – N (cycles to failure) curves for ferrous and nonferrous metals and alloys. The S–N curve can generally be described as shown in Fig. 14b as

$$\log N_f = 10\,(1 - S/\sigma_{max}). \tag{13}$$

Considering the elastic and plastic strain components of fatigue life in terms of strain versus N_f

$$\begin{aligned} &\text{Elastic}: \ \Delta\varepsilon_e/2 = \sigma_a/E = \left(\sigma_f'/E\right)(2N_f)^b \\ &\text{Plastic}: \ \Delta\frac{\varepsilon_p}{2} = \frac{\sigma_a}{E} = \varepsilon_f'\left(2N_f'\right)^c \end{aligned} \tag{14}$$

where $\Delta\varepsilon_e/2$ and $\Delta\varepsilon_p/2$ are the elastic and plastic strain amplitudes, respectively, σ_a is as shown in Fig. 14a, σ_f' is the fatigue strength coefficient equal to the stress intercept at $2N_f = 1$ (Fig. 14b), E is the Young's modulus, ε_f' is the ductility coefficient equal to the strain intercept at $2N_f = 1$, and b and c are the fatigue strength exponent fatigue ductility exponent, respectively. Both equations can be appropriately plotted (log–log) to provide a corresponding straight line with slope b or c, respectively. For low total strains the elastic component in Eq. 14 will dominate, while for large total strains the plastic component in Eq. 14 will dominate.

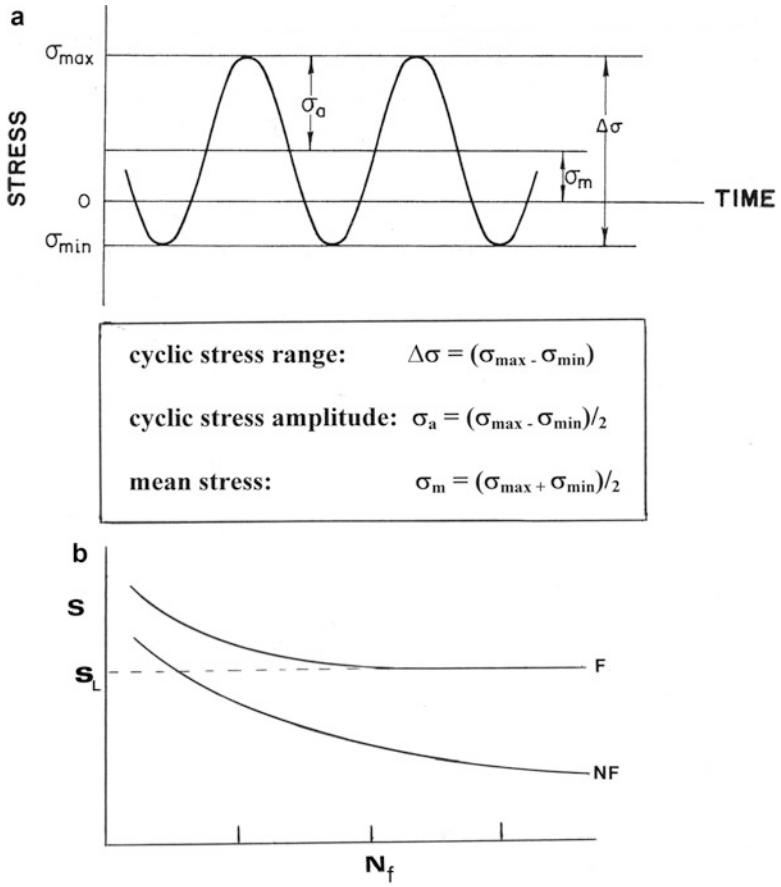


Fig. 14 Fatigue parameters (**a**) and stress (S) – number of cycles to failure (N_f) curves (**b**). F in (**b**) shows an endurance limit (S_L) for ferrous metals. NF represents nonferrous metals and alloys

For steels (ferrous metal shown by F in Fig. 14b), a fatigue endurance level or limit is observed as shown by S_L in Fig. 14b which represents a stress level below which there is no fracture, and the material can be cycled indefinitely. For nonferrous metals and alloys (NF in Fig. 14b), there is no endurance limit. Ideally cracks nucleate by cyclic stressing (or straining), but the mechanism is not a simple reversal of nanotonic stress phenomenon. That is a consequence of the inability to simply reverse plastic shear as implicit in Fig. 1b. Ideally the material can work harden in one direction by dislocation glide, but the reverse direction will create new dislocations in that direction which can interact with those in the initial part of the cycle. As this continues, or as the cycling frequency is changed, more complex interactions will ensue. For very low-cycle fatigue, one might envision reversed creep, while for very high-cycle fatigue, the situation is more chaotic; and diffusion would not play a role. In ductile fcc metals, for example, a simple slip plane or

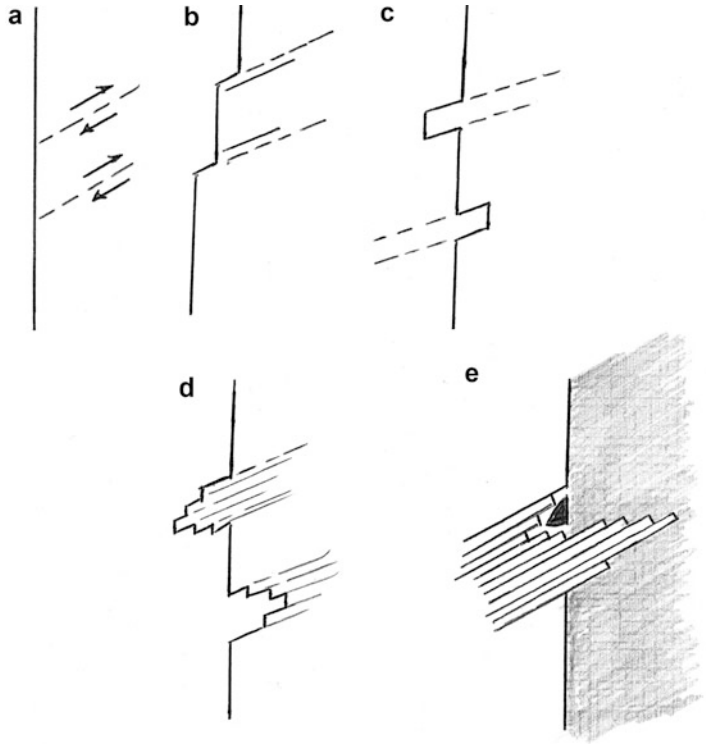


Fig. 15 Fatigue crack formation by slip band/shear reversal at an interface (A to C). D and E show details of crack nucleus development. The shaded region in E can represent a free surface where intrusion–extrusion surface features form. The crack is indicated by dark wedge in the intrusion region in E

reverse glide mechanism has been used to illustrate crack nucleation, especially involving an interface which can be correspondingly sheared, including of course the free surface. This is illustrated in Fig. 15 as an intrusion–extrusion model, depending upon perspective, and could operate in dislocation cell walls (Fig. 2), at grain or phase boundaries, and at free surfaces where slip bands meet the surface. Such regular, cyclic features might be expected to uniquely define the fatigue fracture surface in contrast to other modes of fracture, and it will be observed later that this is indeed the case. It should, however, be mentioned that although the intrusion–extrusion model or concept has been proposed decades ago, it has not specifically been identified as a fatigue crack precursor (Sanjid 2012).

Fatigue crack nucleation, perhaps in some form characteristic of Fig. 15, can lead to coalescence and growth similar to monotonic fracture implicit in Figs. 6 and 7. A simple expression for crack growth under cyclic loading conditions can be written as

$$da/dN = C\,(\Delta K)^m \tag{15}$$

where a is the crack length and N is the number of cycles, C and m are constants, and ΔK s; K_{max}–K_{min}, where K_{max} and K_{min} refer to the maximum and minimum stress intensity factors corresponding to the maximum and minimum loads (or stresses), respectively, and as generally expressed in Eq. 3. However, Eq. 15 is also significantly influenced by temperature as well as the cyclic frequency or rate of load (stress) reversal.

While it is implicit from Fig. 11 that deformation mechanisms dictate fracture mechanisms to a large extent, it is also perhaps implicit that these mechanisms can have an observable influence on both deformation microstructures and subsequent facture, or fracture surface structure. Figure 16 provides some evidence for this on comparing deformation mode-related microstructures in 316 stainless steel, a low-stacking fault free energy fcc alloy. As expected (see chapters "▶ Line Defects: Dislocations in Crystalline Materials" and "▶ Planar Defects: Crystal Interfaces"), the monotonic (tensile) deformation shows linear dislocation and stacking fault arrays in Fig. 16a, b as the strain increases from 0.2 % to 2 %, respectively, at a strain rate of 10^{-3} s^{-1}. Correspondingly, for explosive shock (compression) loading at a corresponding strain rate of $\sim 10^{6}$ s^{-1}, dense twin-fault arrays on corresponding (and intersecting) {111} planes are observed. In contrast to linear dislocation and twin-fault arrays in Fig. 16a–c, Fig. 16d shows dislocation cell structures formed in association with a 316 stainless steel sample strained to fracture at a (creep) strain rate of 10^{-5} s^{-1} at 200 °C. Similarly, Fig. 16e shows dislocation cells formed in cold-drawn and fatigued 316 stainless steel. The 316 stainless steel sample in Fig. 16e was cycled to failure at a rate of ~180 cycles/h, a low-cycle fatigue rate. There is a notable difference in both the cell size and cell wall thickness and structure when comparing Fig. 16d for creep straining versus Fig. 16e for cyclic (fatigue) straining. These modes of deformation allowed for more equilibrium (lower energy) cell structures to form in contrast to linear defect arrays formed in Fig. 16a–c involving tensile straining and shock-impact at high pressure.

Examination of Figs. 1–16 in retrospect illustrates the complexity of fracture and fracture modes even in crystalline or polycrystalline metals and alloys. However, this complexity is even more exaggerated when diffusional effects illustrated in Fig. 12 allow for embrittling elements to migrate to grain boundaries, such as hydrogen, oxygen, etc.; and surface cracks are compromised by environmental reactions promoting *stress-corrosion cracking, corrosion fatigue*, and the like. Temperature also has a significant effect, even promoting ductile–brittle transitional behavior in steels and other materials, including polymers and other nonmetallic materials. Many of these features are apparent in the appearance of fracture surface morphologies.

Comparative Fractography of Materials: Fracture Mode Observations and Performance Features

In contrast to Fig. 2 illustrating microstructures contributing primarily to ductile fracture, especially in metals, Fig. 17 shows schematic classifications of fracture morphologies and associated fracture processes for metals, ceramics, polymers, and

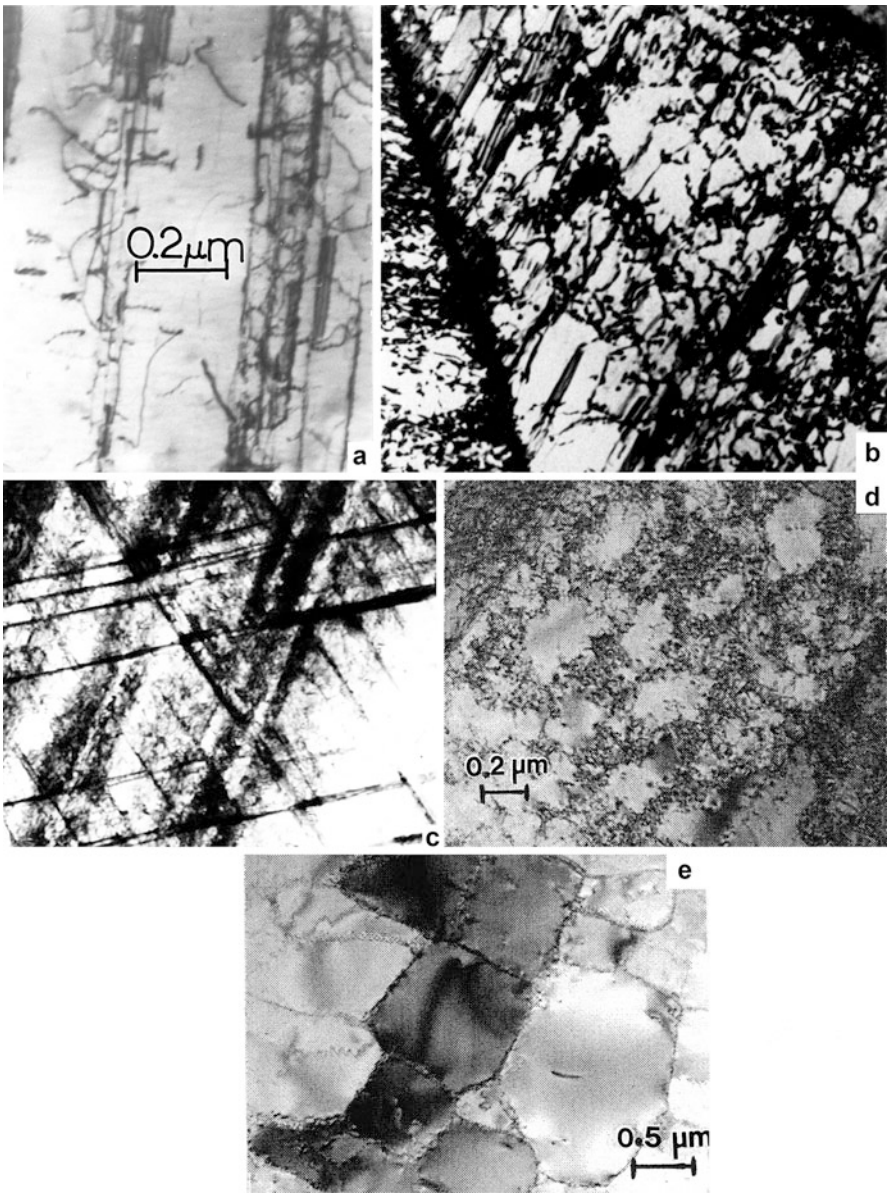


Fig. 16 TEM images for deformation/fracture mode variations in 316 stainless steel. (**a**) Tensile strain to 0.2 %. (**b**) Tensile strain to 2 %. (**c**) Shock loading at 35 GPa. (Surface orientation is (111)). (**d**) Fatigue fracture at 200 °C. (**e**) Creep (10^{-5}/s) fracture at 650 °C

directional composite materials. The fracture processes depicted in Fig. 17 are of course influenced by the fracture mode: tension, compression, shear, impact, creep or fatigue, as well as temperature and the operating environment. Related performance

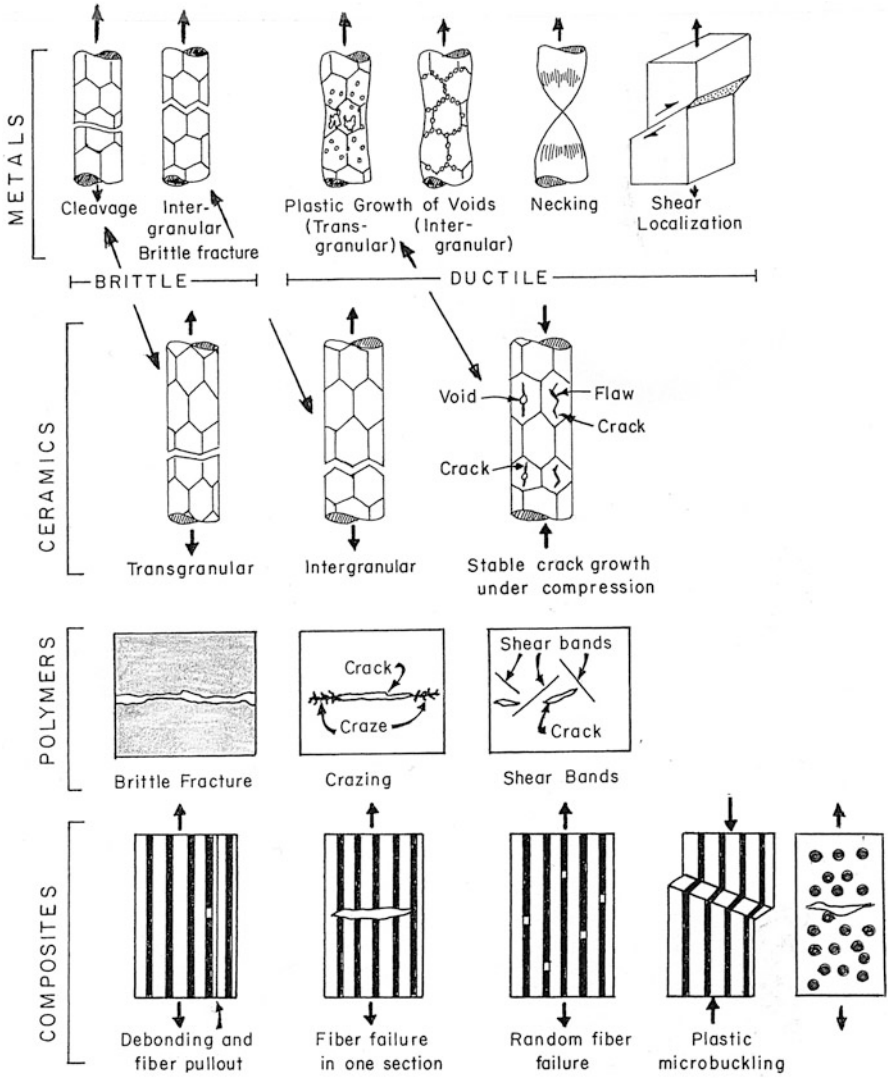


Fig. 17 Fracture morphology and process classification schematics (Adapted from Meyers and Chawla (2002))

relationships and process parameters are illustrated generally in Fig. 18, while Fig. 19 shows the creation of ductile-tensile, ductile-shear, and brittle tensile or impact-related fracture surfaces and surface features. It should be noted that the specific fracture or fracture mode issues outlined schematically in Fig. 18 will be referred to later in discussing each (Hull 1999).

As implied in Fig. 2, any heterogeneity in a material can nucleate cracks through the production of a stress concentration or strain localization. In most engineering

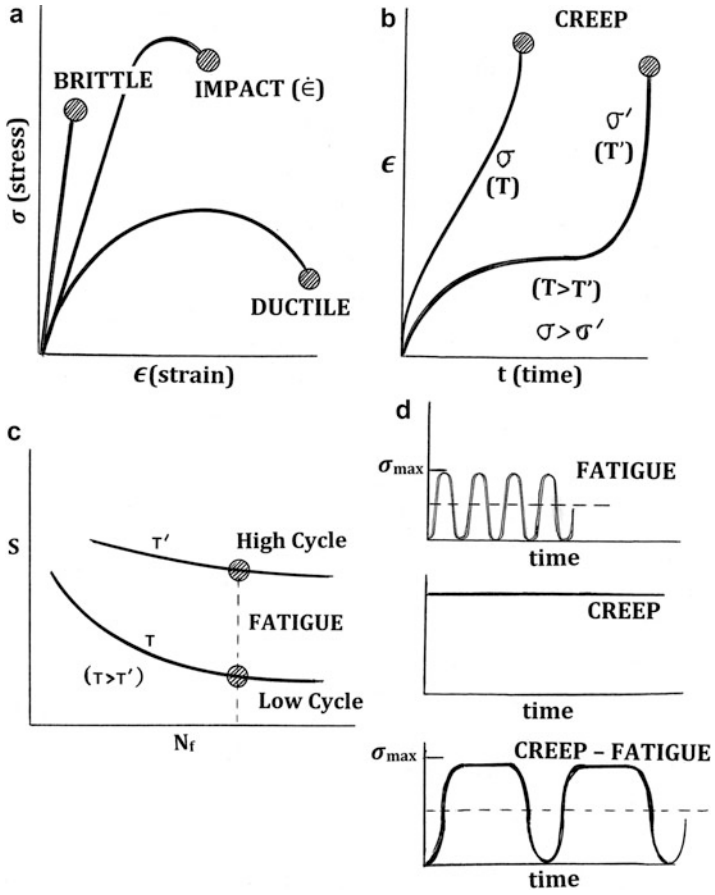


Fig. 18 Fracture mode performance characteristics/parameters. (**a**) Stress–strain diagrams at ~ room temperature. Impact also implies high strain rate ($\dot{\epsilon}$) deformation. (**b**) Creep (ε−t) at high and low stress (σ) and temperature (T). (**c**) Fatigue characterized by stress (S) versus number of cycles to failure (N_f). Temperature and fatigue cycles are illustrated generally. (**d**) Comparison of fatigue, creep, and creep-fatigue modes: D_{max} versus time

materials, particularly metals and alloys, there are significant numbers of impurities (substitutional or second-phase) where microvoids, microcavities, or microcracks originate as shown in Fig. 3. These can coalesce around the particle (which can be as small as 1–2 nm or as large as 1–2 μm). These coalescing microcavities can elongate relative to the applied stress forming egg-shaped and coalescing regions shown in Fig. 19a, with the nucleating particle occupying one end or the other of these separating ellipsoids which form a so-called cup-and-cone fracture surface. Figure 20 shows a fracture surface for a native copper sample failed in tension at a strain rate of 10^{-3} s^{-1}. The magnified image in Fig. 20b shows copper oxide (Cu_2O cuprite) polyhedral particles in the ductile-dimple cups which are spaced roughly 1–2 μm (cup centers). These features are even more compelling in the more

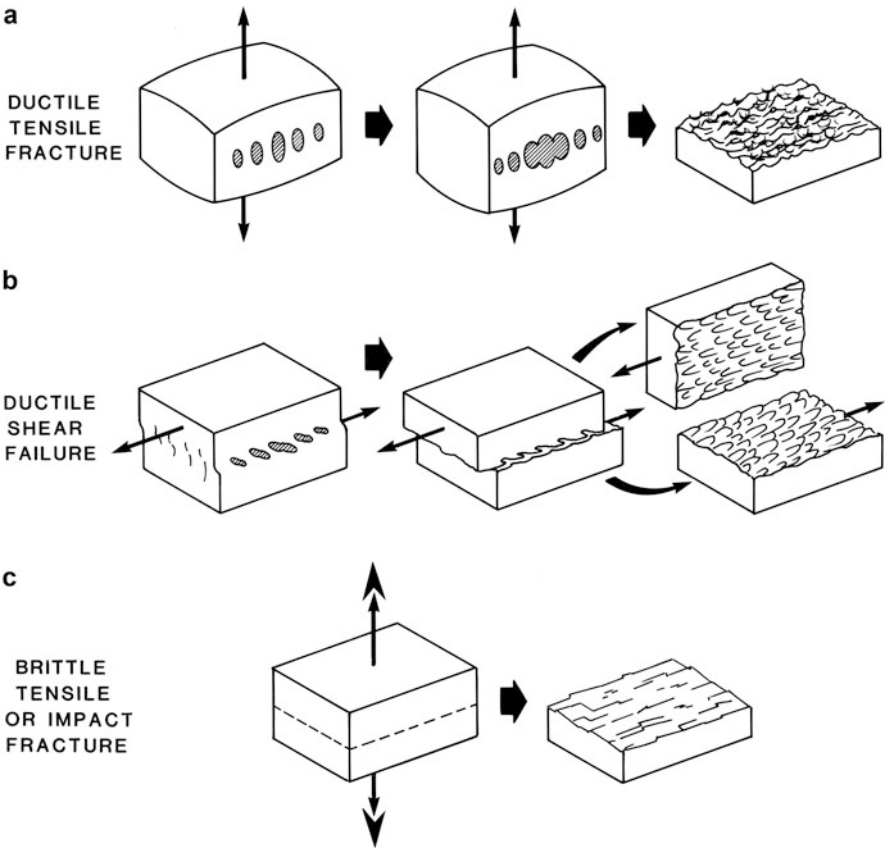


Fig. 19 Schematic views representing common fracture features. (**a**) Ductile-tensile fracture. (**b**) Ductile-shear fracture. (**c**) Brittle-cleavage facture (After Murr (1987))

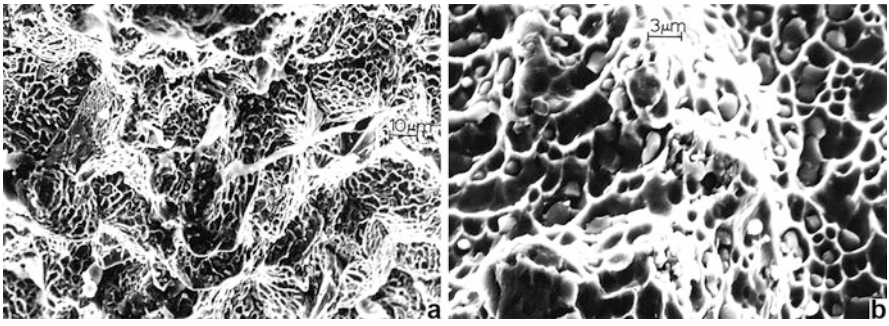


Fig. 20 SEM images for ductile fracture surface of native copper failed in tension. Native copper contains copper oxide (Cu_2O) single-crystal inclusions (precipitates). (**a**) Low magnification, (**b**) higher magnification

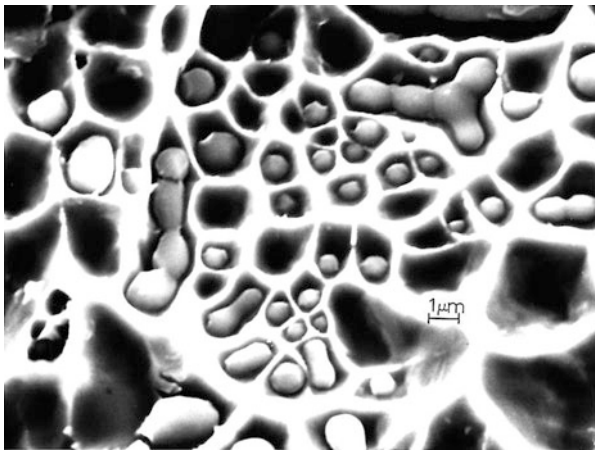


Fig. 21 Magnified SEM image showing copper oxide inclusions in ductile-dimple fracture cups in native copper (Fig. 20)

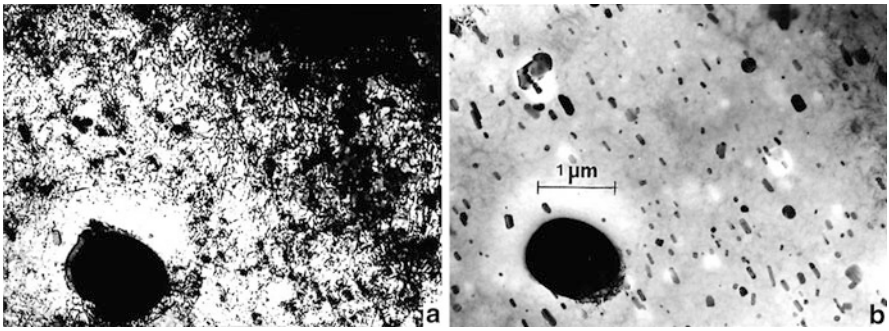


Fig. 22 TEM images for precipitates and second-phase particles in heavily deformed, high dislocation density Al-2024 alloy. (**a**) High dislocation density microstructure. (**b**) Area in (a) tilted for no dislocation contrast showing precipitates and inclusions

magnified SEM fracture surface image shown in Fig. 21, where the Cu_2O inclusions are observed to vary in size from ~300 nm to ~1.2 μm. These inclusions are generally random, although some are aggregated. Second-phase particles of the size and distribution shown in Fig. 21 can have a significant effect on ductile crack propagation and the overall ductility, which is also dependent upon the ease and contribution of matrix dislocation glide.

Figures 22 shows varying sizes of precipitates and second-phase particles in a heavily deformed dislocation matrix in an aluminum alloy. This illustrates the complex role played by microstructure not only in determining strength and other related mechanical properties, but also fracture behavior.

Figures 23, 24, and 25 provide additional evidence for the role of inclusions in ductile-dimple (cup-and-cone) development, their spatial (geometrical) significance, and variances of ductile fracture features when such inclusions systematically vanish. Note the relatively square shape for the ductile cups in the fracture

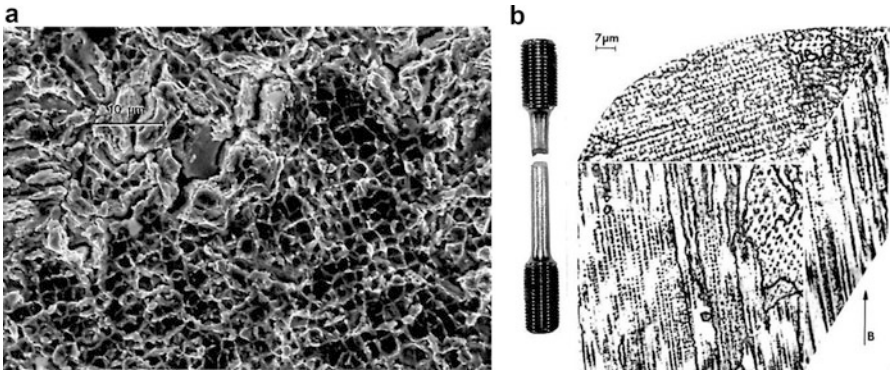


Fig. 23 Ductile fracture (**a**) associated with EBM-fabricated Co-base superalloy (Co-27 Cr-5.5 Mo −3 Fe-1 Ni) specimen fracture in tension (**b**). The fabricated structure is characterized by regular columns of Cr_{23} C_6 cubic precipitates spaced ~2 μm (**c**). The build direction is noted by B in (**c**)

surface structure in Fig. 23a. In addition, the cups themselves are a square array with a cup-center dimension of ~2 μm. This fracture structure corresponds to Co-base superalloy rods fabricated by electron beam melting (EBM) as described in chapter "► 3D and Multidimensional Materials Science" of Part XI, where the scanning beam produced melt zones in an orthogonal (x–y) array spaced ~2 μm. As shown in Fig. 23c, the directional growth within these orthogonal (square) melt zones produced columns of cubic Cr_{23} C_6 precipitates. Upon fracture of this columnar precipitate array (Fig. 23b), the precipitate columnar architecture produced the organized cup-cone (dimple) fracture surface features shown in Fig. 23a. Careful examination of many fracture cups in Fig. 23a will reveal the cubic precipitates in the cup base similar to Figs. 20 and 21.

Figure 24a shows the columnar $Cr_{23}C_6$ precipitate architecture corresponding to the EBM-fabricated Co-base superalloy in Fig. 23c, while Fig. 24b shows a TEM image of individual cubic $Cr_{23}C_6$ precipitates which measure roughly 100 nm on a side. When this as-fabricated product was HIPed at 900 °C for 4 h, the $Cr_{23}C_6$ cubic precipitates were dissolved, producing generally equiaxed, fcc grain structure containing numerous {111}-related annealing twins as shown in Fig. 24c. Figure 24d shows a TEM image with stacking faults within the grains of Fig. 24c, which have essentially replaced the $Cr_{23}C_6$ precipitates in Fig. 24b.

Finally, Fig. 25 shows the fracture surface for HIPed, Co-base superalloy represented in Fig. 24c and (d). Figure 25a, at a comparable magnification to Fig. 23a, shows irregular cup shapes and random geometrical arrays in contrast to the square arrays for Fig. 23a. The magnified image of Fig. 25b does not show any precipitates or inclusions in the fracture cups or dimples.

As noted above, and as illustrated in Fig. 22, the volume fraction of precipitates or second-phase particles has an effect on ductile crack propagation and correspondingly on ductility, which decreases with an increasing volume fraction. In addition, the grain size can also affect ductility, especially for nanograin metals and alloys where

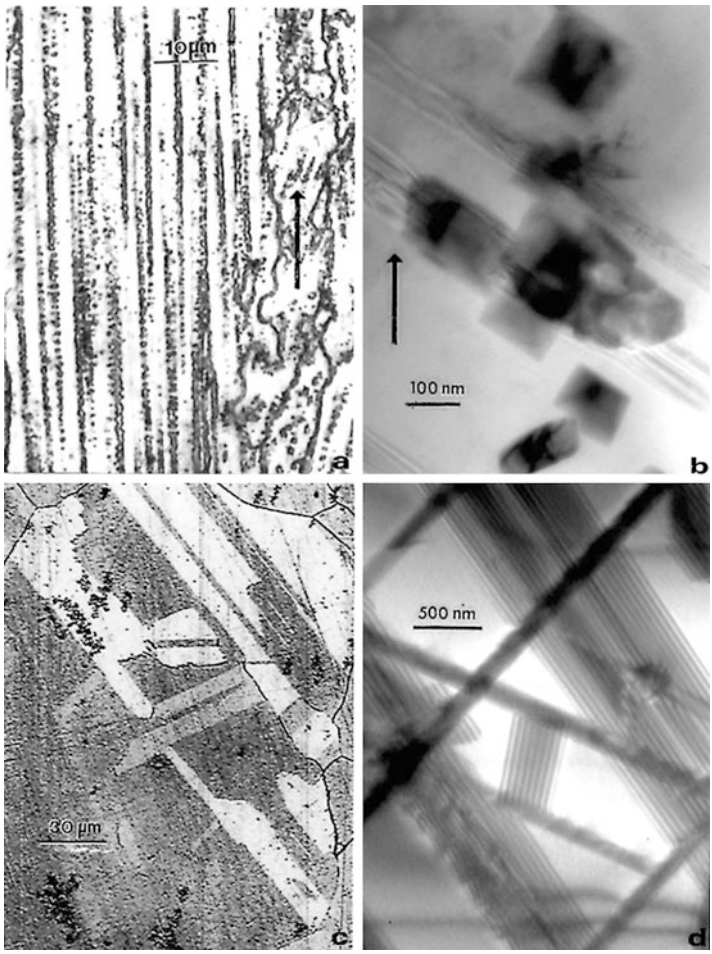


Fig. 24 Comparison of as-fabricated (EBM) Co-base superalloy microstructures ((**a**) and (**b**)) with corresponding, HIPed (annealed) microstructures ((**c**) and (**d**)). (**a**) Optical micrograph showing as-fabricated, columnar $Cr_{23}C_6$ precipitates. (**b**) TEM image of cubic $Cr_{23}C_6$ precipitates in (**a**). (**c**) Grain and twin-grain structure of HIP-annealed component in (**a**). (**d**) TEM image of stacking faults in annealed grains in (**c**)

crack nucleation, coalescence, and propagation will generally increase at some equivalent stress. Small grain sizes usually increase the strength and range of temperatures over which metals and alloys are ductile and exhibit some degree of toughness. In steels, small grain sizes exhibit a smaller *ductile-to-brittle transition temperature*.

As noted in Fig. 17, ductile behavior for metals can involve both intergranular and transgranular void growth, while the ease of dislocation glide can also contribute to necking. For well-annealed metals with no inclusions, necking can become as complete as shown in Fig. 17. In addition, shear failure can also be facilitated as

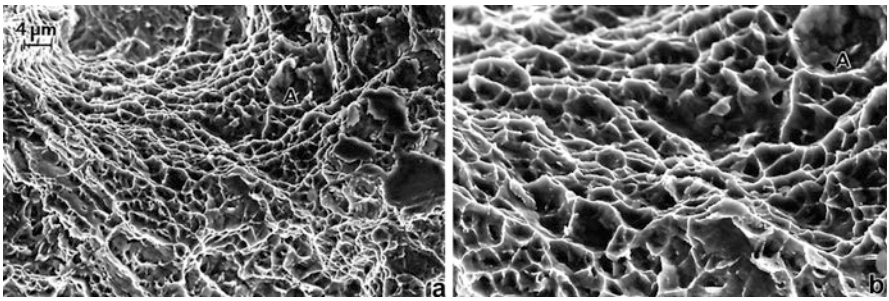


Fig. 25 SEM images of fracture surface for HIP-annealed EBM-fabricated Co-base superalloy components in Fig. 24 (**c**) and (**d**). (**a**) Low-magnification view corresponding to Fig. 23 (**a**). (**b**) Magnified view of (**a**). A indicates a reference area

shown in both Figs. 17 and 19b. Ductile-shear fracture is shown for a crystal section of a Ti-Mo wire in Fig. 26. This wire section represents single-crystal grains aligned in a bamboo structure along the wire length. Grain boundaries perpendicular to the wire axis are shown by arrows in Fig. 26a. Fracture surface shear lips are well defined in Fig. 26d; corresponding to the ideal schematic representation in Fig. 19b.

Brittle fracture, or a ductile-to-brittle fracture transition, occurs with low-temperature deformation and increasing strain rate. This can involve both intergranular and transgranular cracking, while in the case of brittle transgranular fracture, separation can occur catastrophically along specific crystallographic, cleavage planes. However, in amorphous materials, where the effective grain size (D) becomes commensurate with the grain boundary width (Δt) ($\Delta t \sim D$), brittle fracture resembles glass fracture where smooth rupture areas are connected by radiating fracture edges. This is illustrated in Fig. 27. Similar fracture features can be observed in brittle crystalline cleavage where similar fracture edges characterize interconnected cleavage planes as illustrated in Fig. 28 for a Nb–Mo–V–Zr alloy wire which can be compared to the sheared Ti-Mo wire in Fig. 26.

Figure 29 shows similar brittle fracture features for a single-crystal Zn (hcp) specimen fractured at −196 °C. The cleavage plane is (001) and connecting fracture edges connect successive (001) cleavage planes implicit in the hexagonal geometry of the fracture surface. In contrast to transgranular cleavage in Zn in Fig. 29, Fig. 30 shows classic intergranular fracture in polycrystalline Ir (fcc) at room temperature. The grain boundaries are embrittled by oxygen segregation to the interface which drastically reduces the interfacial free energy, and the corresponding adhesive energy.

There are many variances of intergranular (brittle) fracture which can have crack origin in void creep along the boundary, segregation of impurities to boundaries at the time of manufacture which sustain long-term creep stresses, as well as crack enhancement by environmental factors such as corrosion, and especially corrosion at advancing crack tips which accelerate rupture by *stress-corrosion cracking* (*SCC*). These fracture issues are illustrated in examples shown in Fig. 31. In Fig. 31a, intergranular cracks occur heterogeneously throughout the sample,

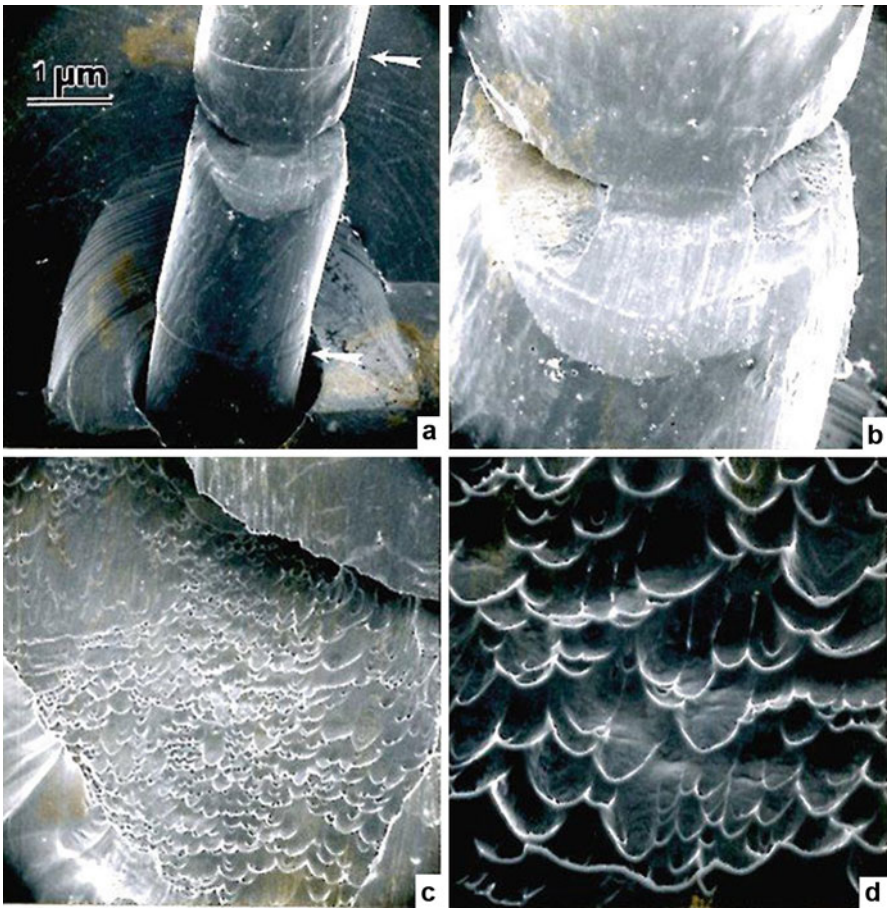


Fig. 26 Ductile shearing of Ti-Mo alloy wire. (**a**) to (**d**) show magnification (increasing) sequence. Note grain boundaries perpendicular to the wire axis at arrows in (**a**) (After Murr 1987)

while Fig. 31b shows a catastrophic primary crack (PC) and secondary intergranular cracks (SC). Figure 31b is typical of pipe steels used extensively in natural gas underground transport. Similarly, Fig. 31c illustrates an extensively cracked natural gas pipeline steel assisted by stress corrosion at the grain boundaries.

Steels and especially steel piping are also characterized by a distinct ductile-to-brittle transition as the temperature is reduced, particularly notable in the context of impact. This is illustrated in Fig. 32 which compares the stress temperature (Fig. 32a) and absorbed energy versus temperature (Fig. 32b) performance curves for a carbon steel. Correspondingly, Fig. 33 shows the ductile-to-brittle transition for HY-80 steel. The contrast in the ductile-dimple fracture surface structure at 0 °C in Fig. 33a is particularly dramatic as a contrast to the classically brittle fracture surface structure shown at −196 °C in Fig. 33d. Note the similarity in the brittle fracture

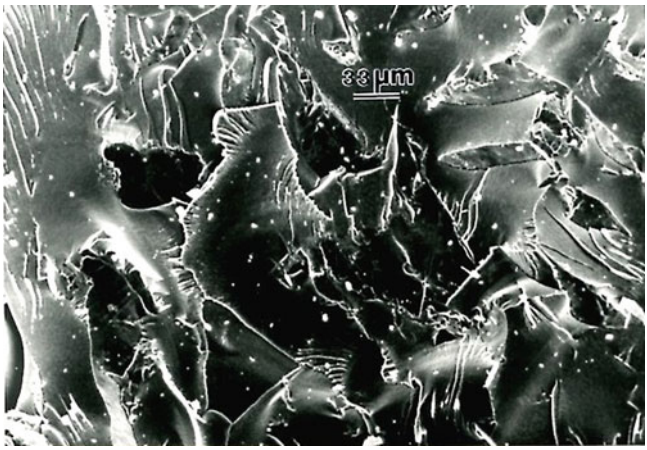


Fig. 27 Brittle fracture surface features for amorphous 80 Fe – 20 B alloy (glassy metal) failed in tension at room temperature

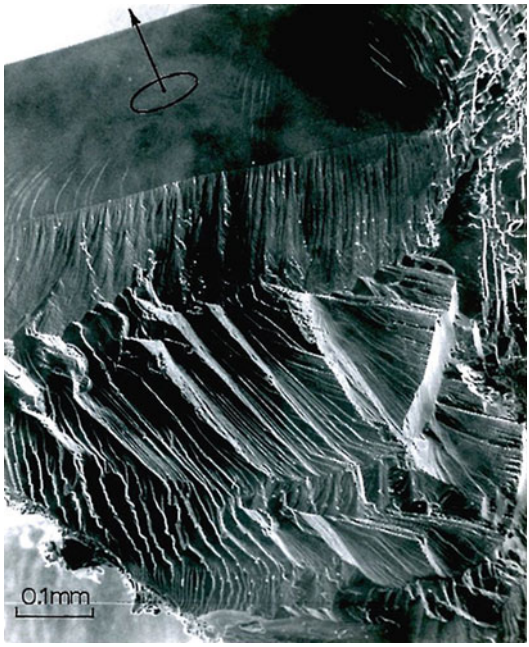


Fig. 28 Brittle-cleavage fracture in Nb-Mo-V-Zr alloy wire failed intension at room temperature (SEM image)

structure of Fig. 33d with the brittle, glassy fracture surface for the amorphous Fe-20B in Fig. 27, structurally or microstructurally quite different metallic alloys.

Although perhaps a little out of the current context, and in some respects a characteristic contrast comparison, it might be of interest to consider melting as an extreme in ductile fracture. It is well documented in ductile fracture, especially at

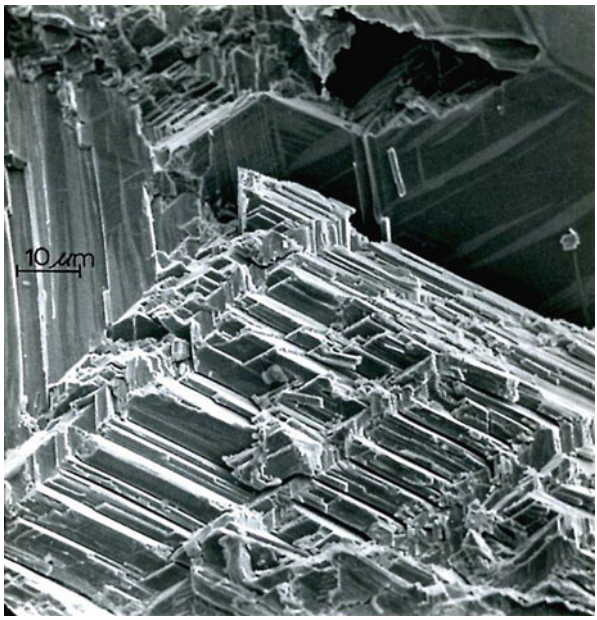


Fig. 29 Cleavage fracture of single-crystal zinc at −196 °C (SEM image)

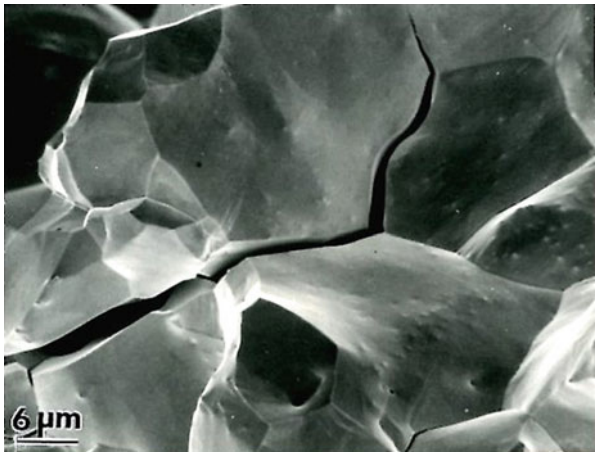


Fig. 30 Intergranular brittle fracture of iridium at room temperature. Uniaxial, tensile failure (SEM image)

high-strain rates, $(\dot{\varepsilon})$ that the fracture surface temperature is elevated (ΔT) at and near the instant of rupture, especially at high strains (ε) since generally

$$\Delta t \propto (\varepsilon)(\dot{\varepsilon}) > T_m. \tag{16}$$

However, when the ductile dimples (or cups) are drawn or stretched to a thickness in the very low nano-dimension range, there will be a suppression of the melting point as illustrated and discussed previously in connection with ▶ Fig. 7 of chapter "3D Printing: Printed Electronics". Since the stretched dimple can represent a significant strain commensurate with a very small thickness, ductile

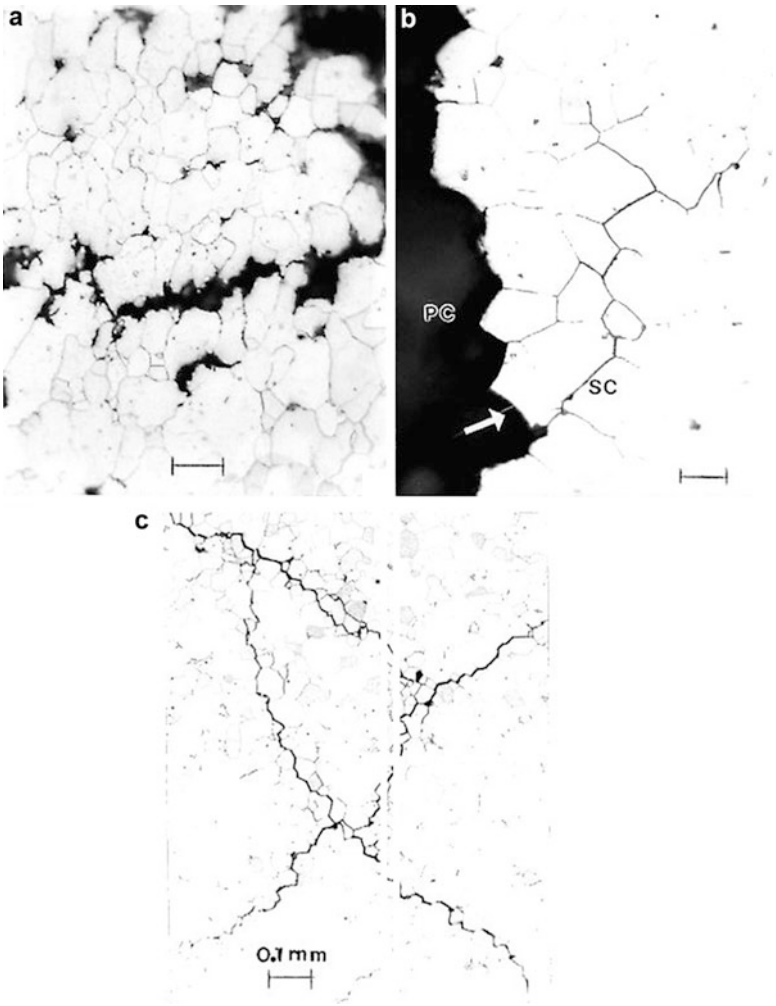


Fig. 31 Examples of intergranular fracture. (**a**) Creep crack initiation and crack propagation in carbon steel pipe. (**b**) Primary (PC) and secondary (SC) cracks in grain interfaces in pipe steel. (**c**) Stress-corrosion cracking in ruptured steel alloy gas pipe line underground (Images courtesy of Dr. Steve Stafford, UTEP)

fracture at high-strain rates can produce melt at the end of the ruptured dimple. This can occur independent of the microstructure or crystal structure and depends strictly upon Eq. 16 and the melting-point suppression. Figure 34 illustrates this phenomenon for a polycrystalline steel sample failed at high-strain rate in contrast to a metallic glass failed at high strain rate. Figure 34c illustrates this phenomenon schematically. Figure 34a, b shows melt beads at the ends of the fracture dimples. Of course this fracture process (Fig. 34a, b) can also be exacerbated by oxidation or other environmental corrosion, especially where melt occurs.

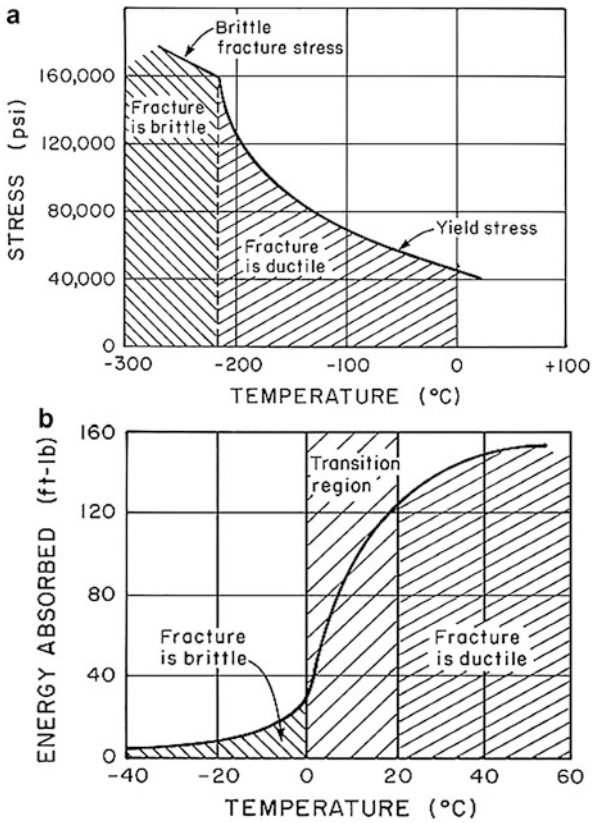


Fig. 32 Ductile–brittle transition in low-carbon steel. (**a**) Tensile test results based upon data of Edlin and Collin (1951). (**b**) Impact test results (Adapted from Murr (1987))

It is worth examining the implications of gas line rupture as illustrated in a few examples in Fig. 31 and to expand such implications to other infrastructure areas where material (especially metals and alloy) failures can have important social and economic consequences. There are of course many hundreds of thousands of miles of pipeline steels in the USA alone. Figure 31 illustrates a few gas line ruptures, but there are also ~55,000 miles of crude oil trunk lines. All of these steels are susceptible over time to corrosion and stress corrosion cracking. One important factor can also include galvanic corrosion and such corrosion catalyzed by soil bacteria. These issues will be discussed in more detail later.

A related transmission infrastructure concern regarding rupture involves more than 500,000 miles of overhead electrical distribution lines carrying high voltages of at least 250 kV, and up to 1 MV, both in the AC or DC modes. In addition, there is another roughly 450,000 miles of feeder lines from the HV transmission systems. Although smaller gauge copper continues to be used in some of these systems, long distance, HV transmission cable is an aluminum alloy surrounding a steel core because of the density difference between Al and Cu. Figure 35 illustrates these features, while Fig. 36 shows both the HV transmission cabling and the local distribution circuit lines. The birds on the lines in Fig. 36b only draw attention to

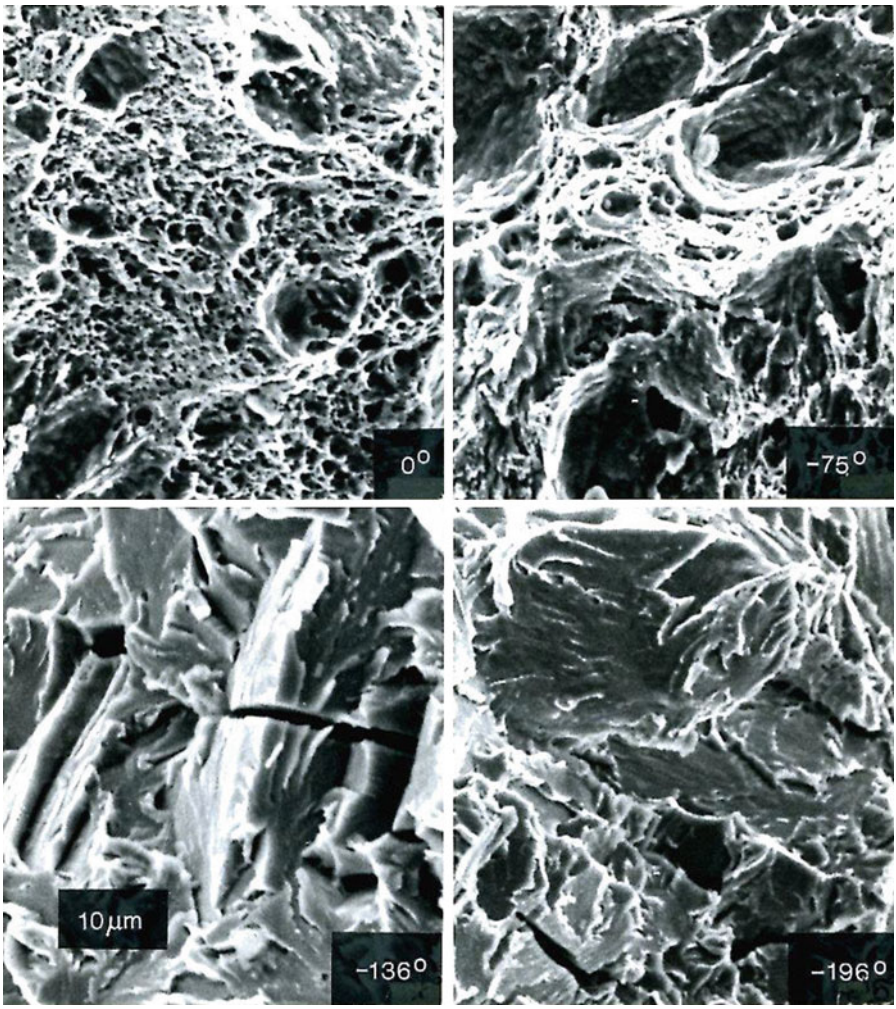


Fig. 33 SEM images for ductile–brittle fractography of HY-80 steel as a function of reducing temperature

rupture prospects, which are even greater for creep in sagging lines from towers shown in Fig. 36a, especially when thick layers of ice form on the cable, or the cables are subjected to high winds. These events can impose large stresses or other modes of deformation, such as fatigue, which can even include thermal cycling in some areas of as much as 30 °C on a daily basis.

There are even more serious rupture concerns in the actual generation of electricity. This involves turbines and their components which not only must sustain high stresses at high temperatures but also resist corrosion in various forms, including corrosion-assisted cracking: stress corrosion cracking. These are very important and very practical materials science and engineering issues.

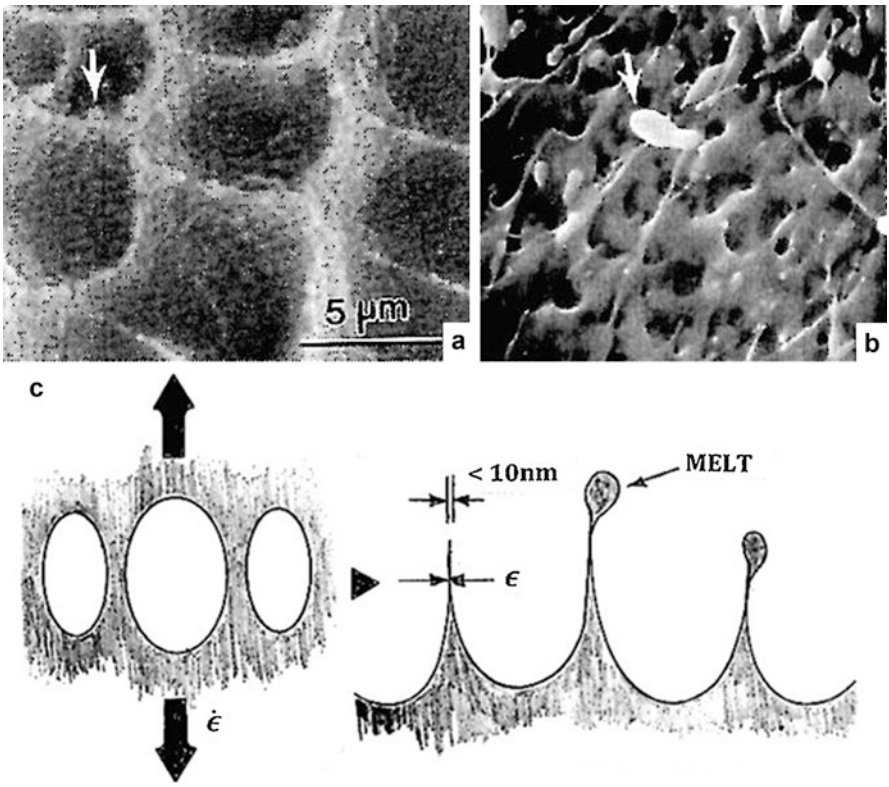


Fig. 34 Melt formation in ductile and ductile-related fracture at high strain rate. (**a**) Melt spheres associated with ductile-dimple walls of a Ti-8Mn alloy failed in high-strain-rate tension (Adapted from Makel and Wilsdorf (1992)). (**b**) Melt drops at fracture surface dimple features for Zr(40) – Ti(14) – Ni (10) – Cu(13) – Be(23) glassy alloy failed in tension (Adapted from Flores and Dauskardt (1999)) (**c**) Schematic views of melt formation in ductile dimple wall thinning at high strain rate tensile failure

Figure 37 illustrates a power plant turbine with its enclosure structure removed. Similar to gas turbine configurations in aircraft, as discussed in chapter "► Examples of Directional Crystal Structures: Gas-Turbine Component Applications in Superalloys" of Part VI, Fig. 37 shows blade sections as large as a person. Certainly the implications of the fracture of just one blade are readily apparent in Fig. 37 in any section of the turbine system. While creep is the obvious fracture contributor, cycle variations and system vibrations can also impose fatigue conditions creating creep-fatigue rupture conditions implicit in Fig. 18d. Two examples of turbine blade fracture are shown in Fig. 38. Figure 38a shows a Japanese aircraft turbine blade fracture in 2011. These blades fractured at the shank area presumably caused by pitting corrosion and stress-corrosion cracking combined with fatigue cracking. Blade fracture caused other engine damage. This engine had accumulated 17,006 flight hours in 3,126 flight cycles, and periodic inspections, including one at 16,752 flight hours, showed no indications of cracking.

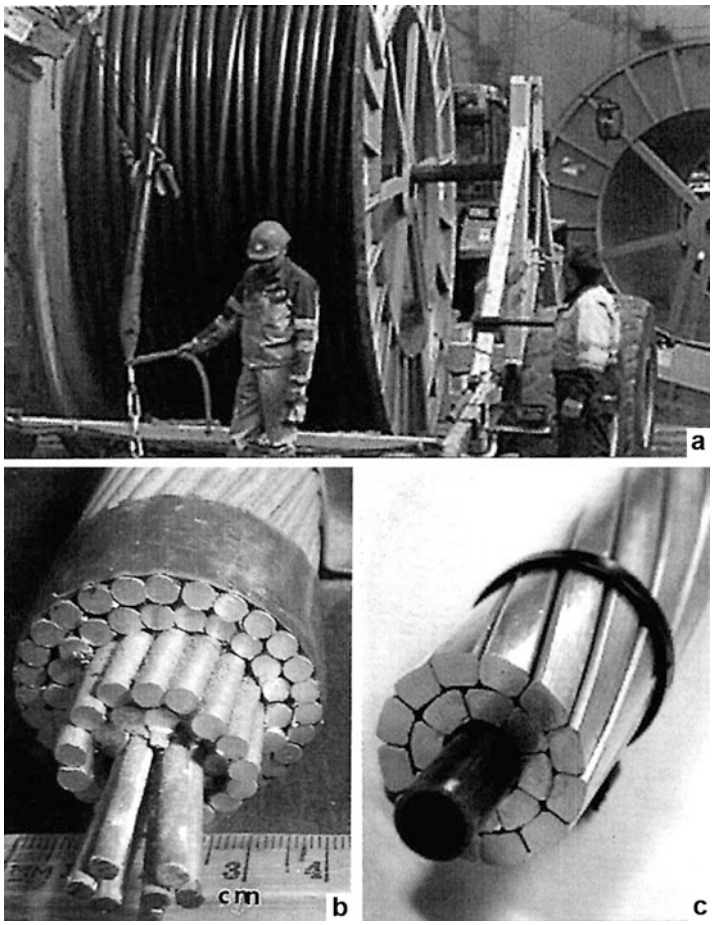


Fig. 35 High-voltage (overhead) aluminum – conductor steel-reinforced (ACSR) cable. (**a**) Polymer-insulator sheathed cable spools. (**b**) and (**c**) show aluminum alloy conductor surrounding steel support cable

Similarly, Fig. 38b shows a larger turbine section of a very large electric generator, similar to Fig. 37, in a Minneapolis, MN-based utility, also fractured at a turbine blade rotor section by stress-corrosion-assisted creep. Blade fracture caused catastrophic damage to the unit which had accumulated a repair cost up to 2013 of more than $146,000,000.

It is interesting and significant that the implications for creep failure of transmission lines shown in Fig. 36b for birds might also be considered in the context of bird strikes on aircraft and damage to turbine systems as well. Large birds, such as geese weighing up to 6 kg, can cause turbine blade damage similar to that shown in Fig. 38a. A bird of this weight would strike the forward turbine section at more than 50,000 ft-lb, enough to do catastrophic damage.

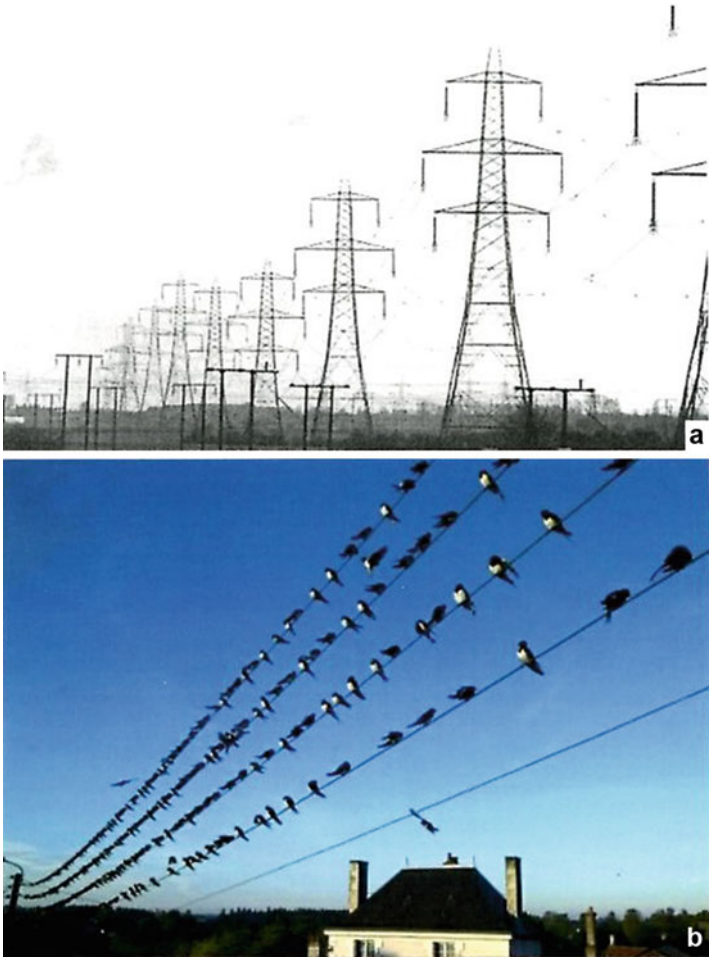


Fig. 36 Overhead electric power transmission systems. (**a**) High-voltage transmission towers with cable sag design using cable as shown in Fig. 35. (**b**) Local area electricity distribution lines which are often copper conductors. Line loads with birds or ice can lead to creep rupture

Fatigue fracture surface structure or fractography is typically illustrated in Fig. 39.

The smaller or more closely spaced striations in Fig. 39a (∼8/μm) are characteristic of high-cycle (high-frequency) fatigue illustrated qualitatively in Fig. 18c. Correspondingly, the larger striations in Fig. 39b (∼2/μm) are characteristic of low-cycle fatigue shown in Fig. 18c. Figure 40b illustrates the S–N data specific to fatigue fracture of 2014-T6 Al alloy shown in Fig. 39a. The maximum stress also plays a role in the striation prominence. Figure 40a shows the S–N data for 1047 steel which exhibits the endurance limit feature discussed in connection with Fig. 14b. In this regard, it is interesting to observe that many common polymeric

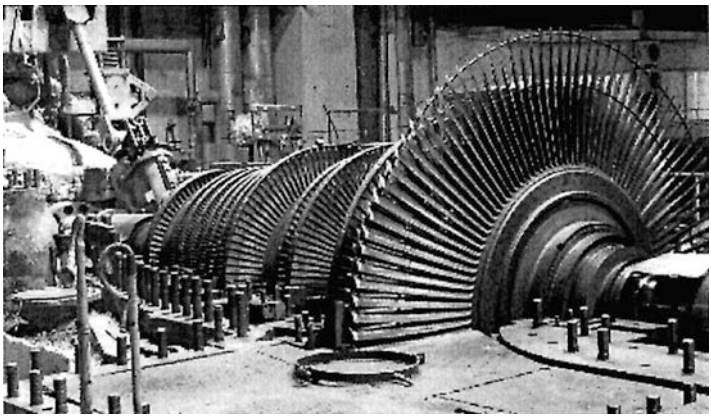


Fig. 37 Exposed steam electric power turbine showing component sections

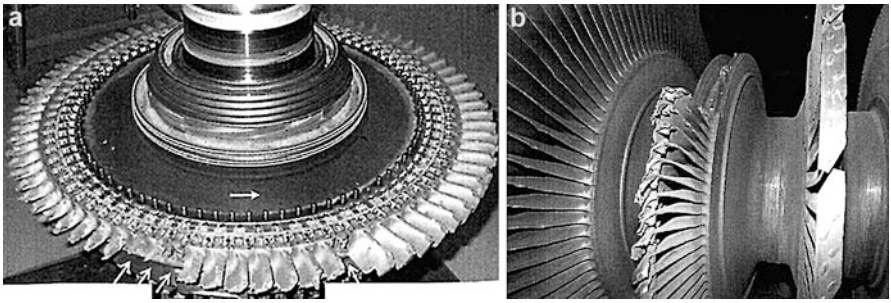


Fig. 38 Turbine blade rupture. (**a**) Nippon (Japan) cargo aircraft engine turbine blade failure (*arrows*). Center arrow (R) shows rotor direction. Fatigue failure aided by corrosion pitting (Courtesy Japan Transportation Safety Board). (**b**) Catastrophic turbine blade fracture by stress-corrosion cracking in Sherco electric power plant generator of Minneapolis-based utility Xcel Energy in 2011 (Courtesy David Shafter, Star Tribune)

materials exhibit this phenomenon as illustrated in Fig. 41a. Figure 41b shows a fatigue or general polymer fracture crack schematic. Regardless of fracture mode, polymer chains become stretched and stretched segments fail.

It might also be noted in connection with Fig. 39a that the fatigue fracture surface is imaged in the TEM using a 2-stage (carbon/metal) replica. Replication imaging has many advantages as described in detail in chapter "► A Brief Introduction to Quantum Mechanics" of Murr (1991). On the one hand, a replica can be made of any surface however large it is. In addition, the same area can be replicated repeatedly. For example, such selected-area replication can be applied to a small stress concentration area or an aircraft landing gear to follow surface crack propagation and growth by simply applying a plastic tape to the designated surface and pulling off, coating with carbon in a graphite arc

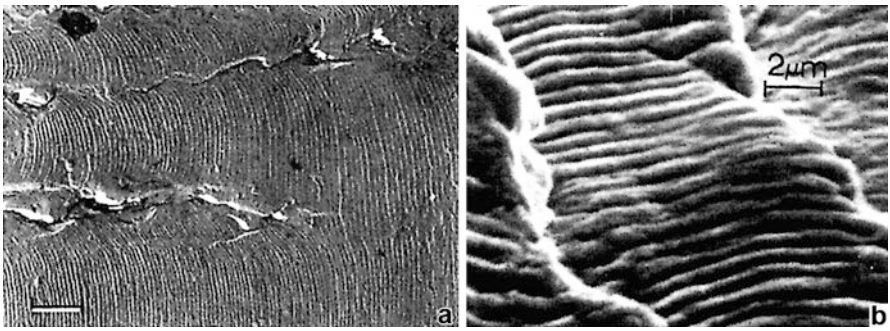


Fig. 39 Fatigue fracture surface striations. (**a**) High-cycle fatigue in Al-2014-T6 alloy observed by replication TEM. (**b**) SEM image showing low-cycle fatigue fracture surface in OFHC copper (Magnification marker in (**a**) is the same as shown in (**b**))

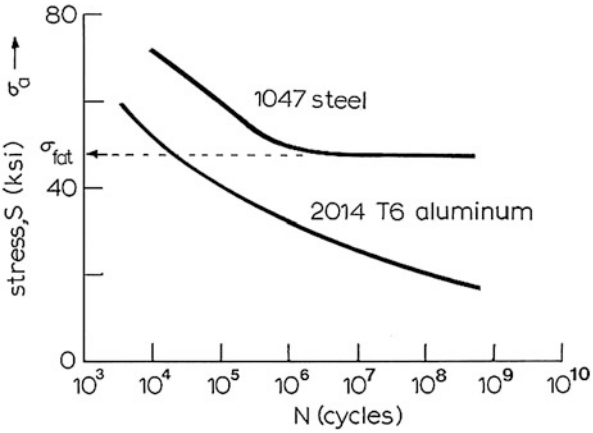


Fig. 40 Stress cycles to failure (S-N_f) curves for 1,047 steel and 2014-T6 aluminum. σ_{fat} is the endurance limit (S_L) as shown in Fig. 14b. Stress in ksi. 1 ksi = 6.9 MPa

(in vacuum), shadowing with a metal to enhance surface–structure contrast, and dissolving the plastic base. Replicas can also be viewed directly in the SEM to produce images which are essentially the same, as illustrated on comparing Fig. 39a, b.

Cracking often occurs within fatigue striations and fracture surface structures such as fatigue striations are often obliterated by rubbing of the partially fractured surface. These features are illustrated in Fig. 42 which shows two views of a small diameter Cu gas line connected to a small stove in a travel trailer which failed by travel-related vibrations and associated fatigue.

Although Fig. 18c illustrates generally qualitative effects of temperature on the S–N fatigue response, Fig. 43a shows more specific data for a 5 Cr – 0.5 Mo steel. Figure 43b shows the effect of grain size variations by an order of magnitude on the fatigue strain amplitude versus the number of cycles to failure (N_f) for 304 stainless steel. Figure 43c, d compare the fatigue crack growth behavior (da/dN) with the stress intensity factor difference (Δk) (as expressed in Eq. 15 with fatigue (cycle) frequency, respectively for an 18 Cr-Nb stainless steel.

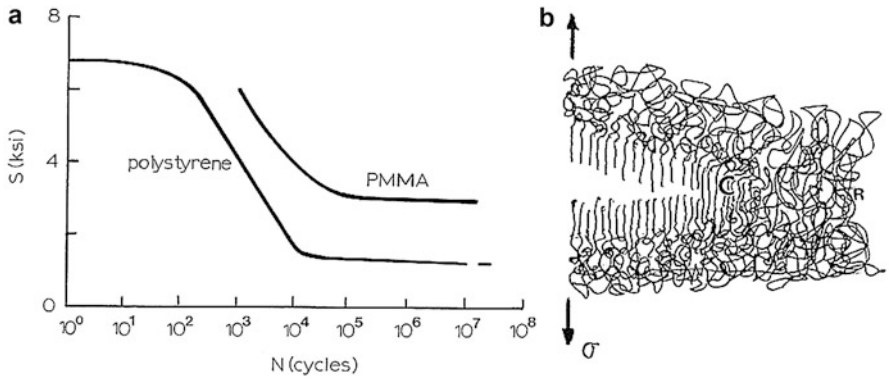


Fig. 41 (**a**) Stress cycles to failure (fatigue) (S-N_f) curves for polymeric materials. Stress in ksi: 1 ksi = 6.9 MPa. (**b**) Polymer crack development and propagation schematic. Locally oriented polymer chains fail by crazing (C) (rupture of polymer chain). R represents the random polymer chain structure

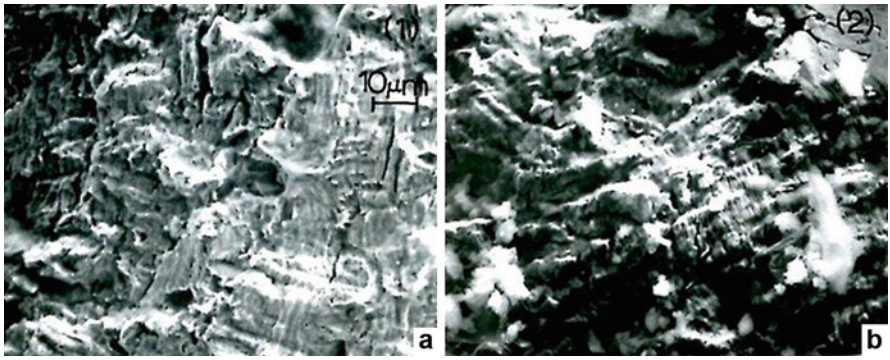


Fig. 42 SEM images showing fatigue fracture surface regions for copper pipe rupture. (**a**) Cracks coincident with fatigue striations. (**b**) Debris and rubbing of fatigue striation surface features

Although failure in fiber-reinforced composite materials can occur by crack initiation at the fiber/matrix interface where strain incompatibilities exist, and propagation between fibers as shown schematically in Fig. 17, the main cause of failure is fiber pullout. Catastrophic crack development and propagation occurs along the fiber/matrix interface as a consequence of weak bonding, leading to delamination. The work for such debonding is given generally by

$$W_D = \pi d^2 \sigma_f{}^2 \ell_D / 24 E_f, \tag{17}$$

where d is the fiber diameter, σ_f is the fiber failure strength, ℓ_D is the length of the debonded zone along the fiber (fiber/matrix interface), and E_f is the fiber Young's modulus.

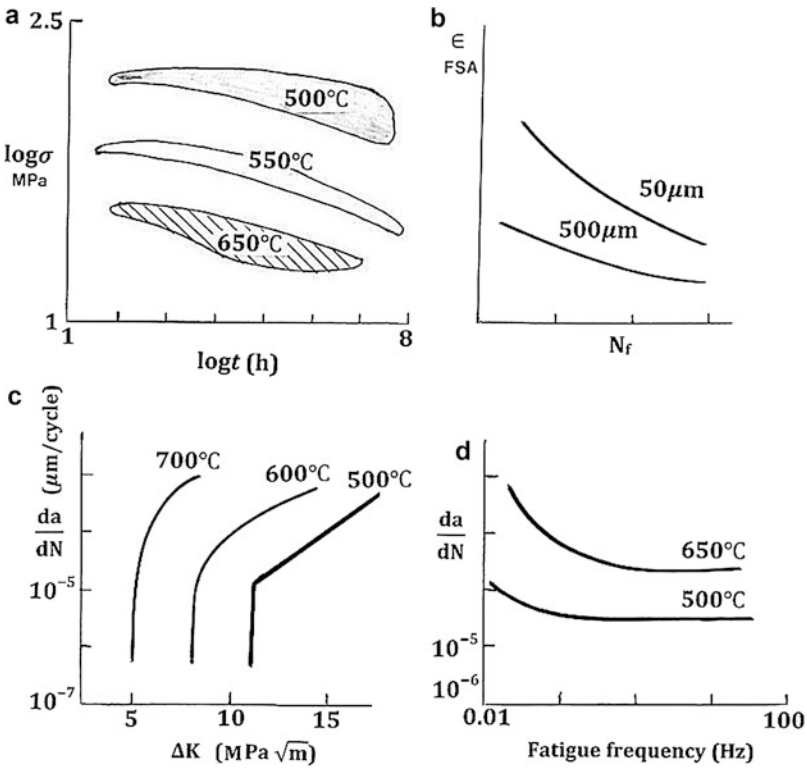


Fig. 43 Fatigue performance features for select commercial alloy steels. (**a**) Temperature effects in fatigue of 5Cr-0.5 Mo steel. (**b**) Grain size effects on fatigue strain versus cycles to failure for 304 stainless steel at 900 K. (**c**) Fatigue crack behavior (growth rate) at temperature for 18 Cr-Nb stainless steel. (**d**) Temperature effects on fatigue crack growth rate (da/dN) versus fatigue cycle frequency (Hz) for (**c**)

Figure 44a illustrates this phenomenon for ceramic SiC fibers in a SiC matrix, while Fig. 44b shows plant fibers (polymeric phormium or flax leaf fibers) in a polylactic acid (PLA) matrix. Brittle fracture of the ceramic SiC fibers in Fig. 44a is in contrast to ductile-like fiber fracture of the flax fibers in Fig. 44b.

Finally, Fig. 45 illustrates a fairly general example of crack nucleation (initiation) and growth fractography. The example shows a fatigue crack initiation site in an Al-7xxx series aircraft airfoil frame structural member in Fig. 45a, while Fig. 45b–d shows enlarged SEM views of the fracture surface structure. Figure 45b shows a section at B in Fig. 45a with a mixture of fatigue striations (arrow) and ductile-dimple rupture zones, while Fig. 45c shows a region marked C in Fig. 45a where fatigue striations are mixed with ruptured fracture features. Figure 45d illustrates the fracture features in the catastrophic rupture region of the propagating crack where fatigue striations have vanished. These crack initiation

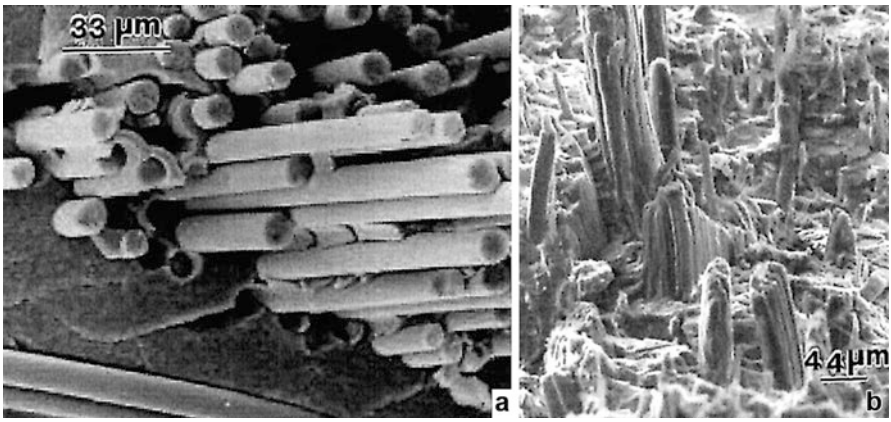


Fig. 44 Failure of fiber-reinforced composite materials. (**a**) Pullout and brittle fracture of SiC fibers in SiC matrix. (Wikipedia). (**b**) Pullout and ductile fracture of flax fibers in PLA polymer matrix (Courtesy Carlo Santulli, Sapienza University of Rome. (**a**) and (**b**) are SEM images)

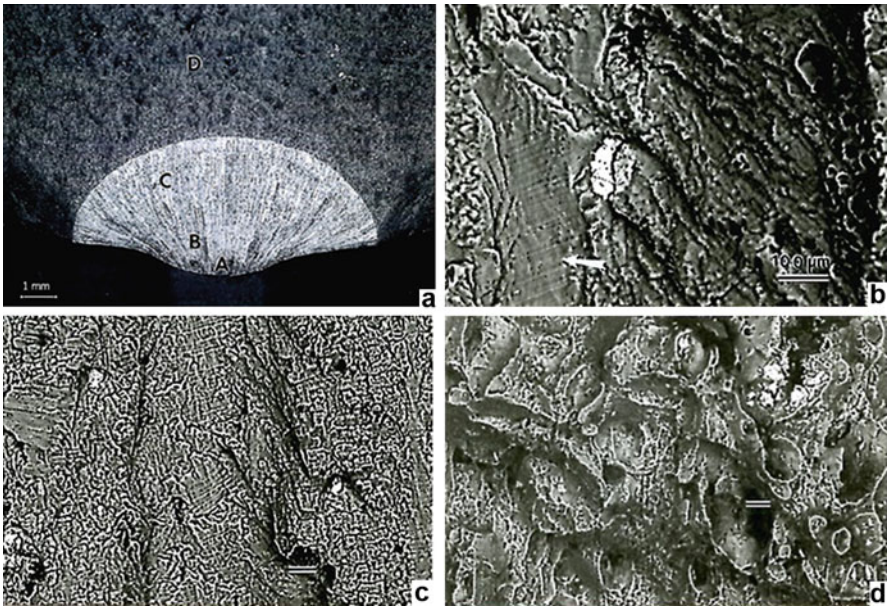


Fig. 45 Fracture initiation and catastrophic crack growth for fatigue failure of aluminum alloy airfoil frame structure. (**a**) Low-magnification light metallograph view of the crack initiation zone (A), crack propagation, and catastrophic rupture. (**b**) SEM views of fracture surface structure at B. (**c**) SEM view of fracture surface structure indicated at C in (**a**). (**d**) Catastrophic fracture structure zone near D in (**a**) (Courtesy of Prof. Steve Stafford and Matthew Garcia, University of Texas at El Paso). Note all markers are 100 µm in (**b**) to (**d**)

and propagation zones are observed generally in all modes of fracture but with differences in the fracture surface structure or structure/microstructure characteristics specific to the fracture mode or mode mixing.

Strain localization in the many forms outlined in this chapter, as these relate to fundamental dislocation mechanisms, initiates cracks in the various fracture modes: impact, creep, fatigue, creep-fatigue, and many complex combinations, including very irregular fatigue superimposed on many structural systems. The reader can peruse the many examples of fracture and fracture criteria presented in this chapter to gain an appreciation for both the complexity of fracture and diagnostic approaches for identifying and understanding fracture modes and underlying structural degradation and failure.

References

Armstrong RW (2012) Grain boundary structural influences on nanopolycrystal strength and strain rate sensitivity. Emerg Mater Res S1:31–37

Ashby MF (1983) Mechanisms of deformation and fracture. In: Hutchinson JW, Wu T-Y (eds) Advances in applied mechanics, vol 23. Academic, New York, pp 118–179

Edlin AS, Collin SC (1951) Ductile-brittle transition in low-carbon steel. J Appl Phys 22:1296–1300

Flores KM, Dauskardt RH (1999) Local heating associated with crack tip plasticity in Zr-Ti-Ni-Cu-Be bulk amorphous metals. J Mater Res 14:638–652

Hull D (1999) Fractography: observing, measuring and interpreting fracture surface topography. Cambridge University Press, Cambridge, UK

Makel DD, Wilsdorf HGF (1992) Localized melting during the separation of high strength tensile samples. Chap. 63. In: Meyers MA, Murr LE, Staudhammer KP (eds) Shock-wave and high-strain-rate phenomena in materials. Marcel Dekker, New York, pp 693–701

Meyers MA, Chawla KK (2002) Mechanical behavior of materials. Upper Saddle River, Prentice Hall

Murr LE (1975) Interfacial phenomena in metals and alloys. Addison-Wesley, Reading

Murr LE (1982) Structure and properties of tensile cracks in stainless steel films: insitu high-voltage electron microscope studies. Int J Fracture 20:117–131

Murr LE (1987) What every engineer should know about – material and component failure, failure analysis, and litigation. Marcel Dekker, New York

Murr LE (1991) Electron and ion microscopy and microanalysis: principles and applications. Marcel Dekker, New York

Sanjid MD (2012) The physics of fatigue crack initiation. Int J Fracture 57:58–72

Wilsdorf HGF (1983) The ductile fracture of metals: a microstructural viewpoint. Mater Sci Eng 59:1–39

Chemical and Electrochemical Degradation and Failure of Materials

Contents

Abstract

Degradation and failure in chemical and electrochemical environments are described and illustrated in the context of corrosion as it relates to oxidation and reduction chemistry, electrode potentials of metals, galvanic corrosion, stress corrosion, and hydrogen embrittlement. Electropositive metal precipitation and its role in crevice and pitting corrosion in copper-containing aluminum alloys are described. Corrosion in atmospheric acidic environments especially by steels and cathodic protection of buried steel pipelines and other vessels is discussed. Fundamentals of embrittlement and especially related diffusional issues, grain boundary segregation, and crack nucleation by gas (hydrogen embrittlement)-, fluid (liquid metal embrittlement)-, and solid-phase production such as metal hybrids acting as stress-related crack nucleation sites are illustrated.

Introduction

It is certainly apparent that many materials, and ideally all materials, are degraded with time in the environment, which includes buried materials and ocean submerged materials. Damage to polymeric materials by the combined action of sunlight (hv) and atmospheric oxygen (O_2) or ozone (O_3) is particularly notable. These effects are exacerbated by thermal cycling and erosion in the form of blowing sand. Degradation of structural materials such as cement or marble (metamorphism

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_54

of sedimentary carbonate rocks such as calcite or dolomite ($CaCO_3$)) and many structural steels, zinc-coated metals and alloys, etc., by exposure to acid rain and snow is also notable in the global environment. Acid rain was coined by Robert Smith around 1872. Mild forms can occur by combining CO_2 and water vapor:

$$\left.\begin{array}{c} H_2O \leftrightarrows H^+ + {+}OH^- \\ H_2O + CO_2 \leftrightarrows H_2CO_3 \\ OH + CO_2 \rightleftarrows HCO_3 \\ H_2O + H_2CO \leftrightarrows HCO_3 + HCO_3 + H_3O^+ \\ HCO_3^- + H^{+3} \rightleftarrows H_2CO_3 \end{array}\right\} \quad (1)$$

The more severe forms include sulfuric acid formation from SO_2 resulting from fossil fuel combustion, volcanoes, forest fires, etc., which account for 70 B kg of SO_2 in the atmosphere:

$$\left.\begin{array}{c} SO_2 + H_2O \leftrightarrows SO_2 \cdot H_2O \leftrightarrows H^+ + HSO_3^- \\ SO_2 + OH^- \rightarrow HOSO_2^- \\ HOSO_2^- + O_2 \rightarrow HO_2^- + SO_3 \\ SO_3 + H_2O \rightarrow H_2SO_4 \end{array}\right\} \quad (2)$$

Nitric acid formation can also occur:

$$\left.\begin{array}{c} NO_2 + OH^- \rightarrow HNO_3 \\ H_2O + N_2 + NO_2 \rightarrow HNO_3 + HNO_2 \\ 3\,HNO_2 \leftrightarrows HNO_3 + 2\,NO + H_2O \end{array}\right\} \quad (3)$$

These reactions can also be catalyzed by lightening.

Degassing of sea salt, coal combustion, and chlorocarbons also forms significant global amounts of HCl which is estimated to compose nearly 40 % of global acid rains.

Atmospheric ozone can also have an effect on acid production:

$$\left.\begin{array}{c} NO + O_3 \rightarrow NO_2 + O_2 \\ NO_2 + O_3 \rightarrow NO_3 + O_2 \\ NO_2 + OH \rightarrow HNO_3 \end{array}\right\} \quad (4)$$

$$\left.\begin{array}{c} 3SO_2 + 3H_2O + O_3 \rightarrow 3H_2SO_4 \\ H_2S + O_3 \rightarrow SO_2 + H_2O \end{array}\right\} \quad (5)$$

While carbonic acid (Eq. 1) is a weak acid, it can dissolve limestone, while the more aggressive acids H_2SO_4, HNO_3, and HCl can react with limestone (marble) structures:

$$CaCO_3 + 2H^+ \rightarrow Ca^{2+} + H_2O + CO_2 \quad (6)$$

Figure 1 illustrates the typical effects of acid-rain degradation on limestone.

Fig. 1 Acid-rain deterioration of marble statue (Courtesy of Wikimedia Commons)

Structural, exposed metals and buried metals can react with acid rain and acidic environments, including soil. For example, buried steel or copper pipe can react generally in the form

$$M + 2H^{+} \rightarrow M^{+} + H_2 \tag{7}$$

and cause almost complete degradation (deterioration) as illustrated in Fig. 2. Similar catastrophic corrosion occurs in saltwater environments as illustrated in Fig. 3a, and reinforcing (steel) bars are a principal cause of deterioration and failure of a variety of concrete structures, especially those exposed to seawater environments or dissolved salts as illustrated in Fig. 3b. This is particularly characteristic of structures in a variety of marine environments and those exposed to deicing salts, including highway and bridge structures. As the corrosion products form on the rebar, stresses created cause cracking and spallation of the concrete matrix which is implicit in Fig. 3b.

Electrochemical Corrosion of Materials

While chemical attack can occur in materials as briefly illustrated above, most corrosion in metals and alloys involves electrochemical cells or galvanic cells (or couples). These form so-called half-cell reactions: one involving oxidation and one reduction. Oxidation reactions produce electrons, while reduction reactions

Fig. 2 Heavily corroded underground steel pipe (Courtesy of Utah State University Buried Structures Laboratory)

Fig. 3 Salt-aerosol-related steel structure corrosion. (**a**) Steel ship hull. (**b**) Steel rebar – reinforced concrete structure showing rebar corrosion and concrete spallation due to corrosion product expansion. (Courtesy Concrete Construction. Net). *Arrows* show exposed, corroded rebar

consume electrons. The oxidation-reduction electrochemistry involved in metal oxidation can be expressed as

$$\left.\begin{array}{l}\text{Oxidation partial reaction: } \mathrm{M} \rightarrow \mathrm{M}^{n+} + n\bar{e} \\ \text{Reduction partial reaction: } n\mathrm{O}_2/2 + 2n\bar{e} \rightarrow n\mathrm{O}^{-2}\end{array}\right\} \tag{8}$$

where M is a metal having a valence n+ and $\bar{e}$ is the electron (negative change).

Consequently for aluminum oxidation,

$$4\,Al^{3+} + 3{O_2}^{-2} \rightarrow 2Al_2O_3 \tag{9}$$

Those metals which are more reactive than hydrogen are assigned negative potentials and are *anodic* to hydrogen and are oxidized to form metal ions, and hydrogen ions are reduced to form hydrogen gas:

$$\left.\begin{array}{l} M \rightarrow M^{n+} + n\bar{e}\,(\text{oxidation}) \\ 2H^{+} + 2\bar{e} \rightarrow H_2\,(\text{reduction}) \end{array}\right\} \tag{10}$$

For those metals less reactive than hydrogen, they are assigned positive potentials and are *cathodic* to hydrogen:

$$\left.\begin{array}{l} M^{n+} + n\bar{e} \rightarrow M\,(\text{reduction}) \\ H_2 \rightarrow 2H^{+} + 2\bar{e}\,(\text{oxidation}) \end{array}\right\} \tag{11}$$

Table 1 lists the standard half-cell potentials for some selected metals at room temperature. For a galvanic couple, the electrode which is oxidized is called the anode, and the electrode where reduction takes place is referred to as the cathode. Metal ions and electrons are produced at the anode (Eq. 10), and since the electrons remain in the metal electrode, the anode is assigned negative polarity. At the cathode electrons are consumed (Eq. 10), and correspondingly it is assigned a positive polarity.

Consider the galvanic couple created by joining a copper pipe section to an iron (steel) pipe section. From Table 1, Fe becomes the anode ($E^{\circ} = -0.44$ V), and the Cu becomes the cathode ($E^{\circ} = -0.337$ V):

$$\left.\begin{array}{l} \text{Anode reaction (oxidation): } Fe \rightarrow Fe^{2} + 2\bar{e} \\ \text{Cathode reaction (reduction): } Cu^{2+} + 2\bar{e} \rightarrow Cu \end{array}\right\} \tag{12}$$

The iron has the more negative potential and oxidizes (corrodes). This is illustrated in Fig. 4. For an iron (or steel) pipe or other structure in an acidic environment (pH < 7), the electrochemical equations become

$$\left.\begin{array}{l} Fe \rightarrow Fe^{2+} + 2\bar{e} \\ 2\,H^{+} + 2\bar{e} \rightarrow H_2 \end{array}\right\} \tag{13}$$

The iron (or steel) then corrodes as shown in Figs. 2 and 3. However, if electrons are supplied to the steel (Fe), dissolution of the metal (corrosion) will be suppressed, and the hydrogen evolution increased (Eq. 13). A DC power supply can be connected to the structure to provide the electrons. Alternatively, electrons can be supplied by galvanic coupling of the steel structure to a more anodic metal (more negative potential in Table 1). This is referred to as cathodic protection and is

Table 1 Standard electrode potentials at 25 °C[a]

	Oxidation (corrosion) reaction	Electrode potential, E° (volts vs. standard hydrogen electrode)
More cathodic (less corrosive)	$Au \rightarrow Au^3 + 3e^-$	+ 1.498
	$2H_2O \rightarrow O_2 + 4H^+ + 4e^-$	+ 1.229
	$Pt \rightarrow Pt^2 + 2e^-$	+ 1.200
	$Ag \rightarrow Ag^+ + e^-$	+ 0.799
	$2Hg \rightarrow Hg_2^{2+} + e^-$	+ 0.788
	$Fe^{2+} \rightarrow Fe^3 + e^-$	+ 0.771
	$4(OH)^- \rightarrow O_2 + 2H_2O + 4e^-$	+ 0.401
	$Cu \rightarrow Cu^2 + 2e^-$	+ 0.337
	$Sn^2 \rightarrow Sn^4 + 2e^-$	+ 0.150
	$H_2 2H^+ + 2e^-$	0.000
More anodic (more corrosive)	$Pb \rightarrow Pb^{2+} 2e^-$	− 0.126
	$Sn \rightarrow Sn^{2+} + 2e^-$	− 0.136
	$Ni \rightarrow Ni^2 + 2e^-$	− 0.250
	$Co \rightarrow Co^2 + 2e^-$	− 0.277
	$Cd \rightarrow Cd^{2+} + 2e^-$	− 0.403
	$Fe \rightarrow Fe^{2+} + 2e^-$	− 0.440
	$Cr \rightarrow Cr^{3+} + 3e^-$	− 0.744
	$Zn \rightarrow Zn^{2+} + 2e^-$	− 0.763
	$Al \rightarrow Al^{3+} + 3e^-$	− 1.662
	$Mg \rightarrow Mg^{2+} + 2e^-$	− 2.363
	$Na \rightarrow Na^+ + e^-$	− 2.714

[a]Reactions represent anodic half-cells. The more negative the half-cell reaction, the more anodic the reaction is and the greater the tendency for corrosion or oxidation.

illustrated schematically in Fig. 5. Although Al has a more negative potential than Fe (Table 1), magnesium has been the metal of choice because it has an even more negative potential.

There are numerous situations in metals and other materials where variations in chemistry can cause selected areas or volumes to become anodic relative to the surrounding matrix, or aggregation or contact of electrochemically dissimilar phases can produce galvanic couples. These can include depleted areas such as in Cr depletion in sensitized stainless steel grain boundaries in the formation of $Cr_{23}C_6$ precipitates, for example, or the migration of some species to a crevice or crack

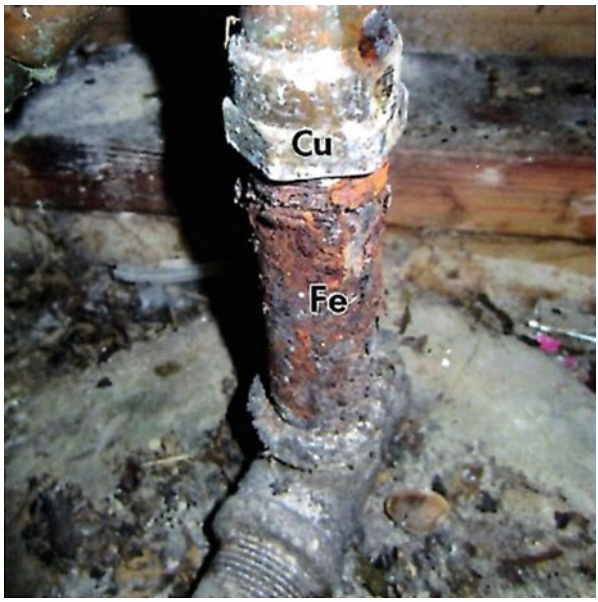


Fig. 4 Junction of copper and galvanized steel water pipes. Both the Zn (galvanized coating) and Fe pipe corrode relative to the Cu (Courtesy of Hergert Inspection, LLC, Seattle, WA)

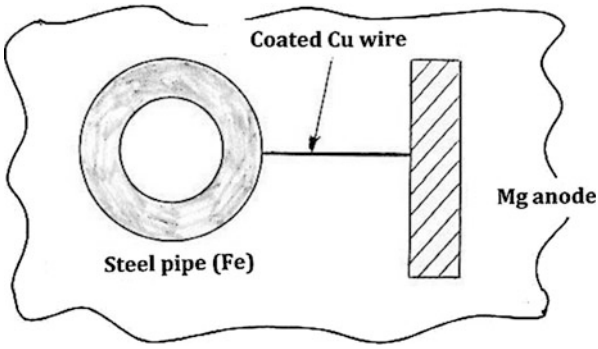


Fig. 5 Cathodic protection schematic. Mg is sacrificial (corroding) metal protecting the Fe

which can promote an electrochemical reaction characteristic of corrosion as illustrated above. Such galvanic contact systems have also been employed in metal extraction along with related electrochemical reactions involving the displacement of a more electropositive metal by a metal ion in a contacting aqueous solution.

For example, Fig. 6 shows the galvanic interaction between pyrite (FeS_2) and chalcopyrite ($CuFeS_2$) in an acid medium containing bacteria. Here the chalcopyrite acts as the anode, while pyrite acts as the cathode. The corresponding reactions become

$$\left.\begin{array}{l}\text{Anode: } CuFeS_2 \rightarrow Cu^{2+} + Fe^{2+} + 2S + 4\bar{e} \\ \text{Cathode: } O_2 + 4H^+ \rightarrow H_2O\end{array}\right\} \tag{14}$$

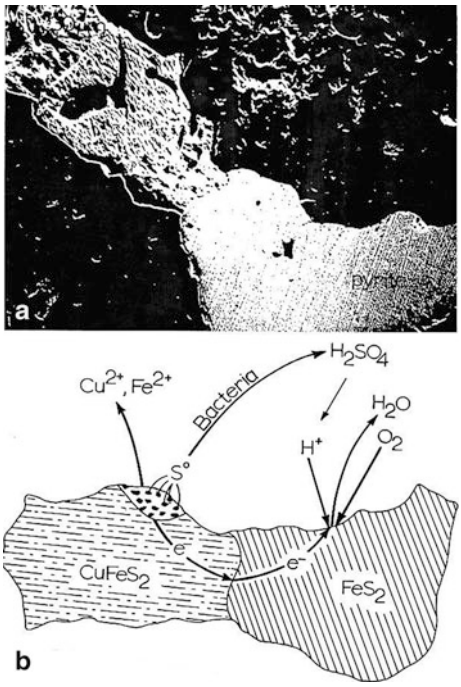


Fig. 6 Contacting chalcopyrite and pyrite inclusions in a quartz matrix creating a galvanic couple which preferentially corrodes the anodic chalcopyrite. (**a**) SEM image. (**b**) Schematic illustration of electrochemical reactions in (**a**) (From Murr 1980)

and the overall galvanic reaction is then

$$\left.\begin{array}{l} CuFeS_2 + O_2 + 4\,H^+ \rightarrow \\ Cu^{2+} + Fe^{2+} + 2S + H_2O \end{array}\right\} \tag{15}$$

The elemental sulfur formed in this reaction can be converted to sulfuric acid:

$$2S + 3O_2 + 2\,H_2O \xrightarrow{bacteria} 2\,H_2SO_4, \tag{15a}$$

where bacteria also catalyze this reaction. It can be noted in Fig. 6 that the anodic chalcopyrite surface is reacted, while the pyrite surface is passivated. The role of bacteria in catalyzing this galvanic reaction is considerable. Figure 7 illustrates this role in a controlled leaching of chalcopyrite waste rock continuing pyrite in galvanic couples shown in Fig. 6, with and without bacterial catalysis in shake flasks, where bacterial catalysis is considerable.

The role of bacteria in other acidic environmental reactions such as the reaction of iron (or steel) underground represented in Eq. 13 can also contribute significantly to pipe or storage fork deterioration. Consequently, pH control can also reduce deterioration and the rate of deterioration of buried structures.

One of the principal recovery technologies for copper from copper sulfate solutions in mine leaching operations prior to about 1985 was a process called *cementation* where iron scrap was used to precipitate the more electropositive

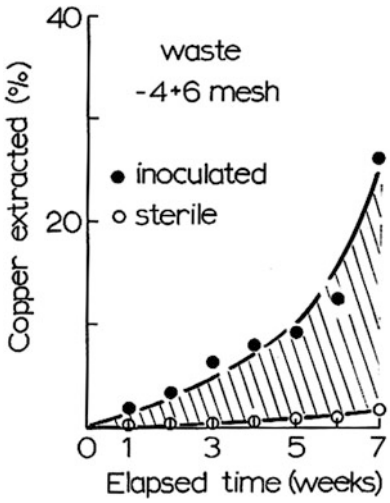


Fig. 7 Leaching of chalcopyrite ($CuFeS_2$) – bearing quartz waste containing chalcopyrite/pyrite galvanic couples as in Fig. 6a ground to size shown and leached in acid-aqueous solutions with (inoculated) and without (sterile) bacteria (*T. ferroxidans*) (pH 2.3, T = 28 °C) (From Murr 1980)

copper in solution (Table 1; Cu(E°) = + 0.337; Fe (E°) = 0.440). The overall cementation reaction is the combination of the respective half-cell reactions:

$$\left.\begin{array}{l} M^{m+} + m\bar{e} \rightleftharpoons M \\ N^{n+} + n\bar{e} \rightleftharpoons N \end{array}\right\}, \tag{16}$$

where with reference to Fig. 8, ions from solution ($M^{mt} = Cu^{2+}$) having a valence m are reduced to elemental metal (M = Cu) on the more electropositive, solid metal substrate (N = Fe). The arrows in Eq. 16 indicate the directions of the reactions. Murr and Annamalai (1978) have shown that aluminum scrap can also provide a cementation regime when the Al_2O_3 oxide layer is removed using Cl^- ion ($\geq$75 ppm minimum in solution). These cementation reactions using iron scrap or aluminum scrap (e.g., cans) can be expressed by

$$\left.\begin{array}{l} Cu^{2+} + Fe = Fe^{2+} + \mathbf{Cu} \\ Cu^{2+} + \frac{2}{3}Al = \frac{2}{3}Al^{3+} + \mathbf{Cu} \end{array}\right\} \tag{17}$$

As shown in Fig. 8, as the copper deposits as a neutral metal precipitate on the Fe or Al substrate, the substrate is dissolved (N^{n+} in Eq. 16) as illustrated in Fig. 9 for an aluminum can. The corresponding nucleation and growth of the Cu from solution is pH and temperature dependent as illustrated for Cu deposits in Fig. 10. Figure 11 illustrates the details of nucleation of Cu from solution onto a thin Al film observed after different contact times shown forming optimal microdendritic growth of Cu as observed in the TEM.

It has been observed that the corrosion of dilute aluminum alloys used in aircraft construction, such as Al-2024 (containing ~4 w/o Cu), and especially older military and commercial aircraft, involves galvanic couples established between copper-rich

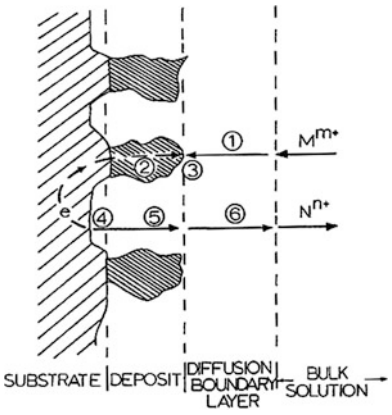


Fig. 8 Schematic showing electropositive copper (M^{m+}) precipitating (cementing) on reacting substrate (Fe or Al: N^{n+}). Cu^{2+} ions (M^{m+}) diffuse to deposit/solution interface region at (*1*). Electrons are conducted from dissolution to deposition sites (*2*), allowing reduction of atoms (M) into deposit (*3*). Substrate (Fe or Al) metal ions (N^{n+}) at dissolution sites are released into solution (*4*) and transferred to deposit/solution interface (*5*) where they diffuse into the bulk solution (*6*) (From Murr and Annamalai 1978)

intermetallic inclusions/precipitates (such as Al_2Cu Mg, Al-Cu-Si, and Al-Cu-Mn-Fe particles) and the adjacent copper-depleted matrix (Obispo et al. 2000). The large precipitate (inclusion) in ▶ Fig. 22 of chapter "Fracture Modes and Mechanisms" illustrates an Al-Cu-Mg second-phase particle typical of the Al-2024 alloy microstructure. In a NaCl (ocean salt) aqueous environment, for example, Cl^- ion breaks down the Al_2O_3 oxide layer exposing the alloy surface and allowing corrosion to dissolve Cu from the matrix and from Cu-containing precipitates of various kinds (▶ Fig. 22 of chapter "Fracture Modes and Mechanisms"), forming a concentrated Cu^{2+} solution near the surface of the Al alloy or Fe-rich precipitates from which the Cu^{2+} will precipitate out as shown schematically in Fig. 9 and in Eq. 17. In general, Al corrosion or *pitting corrosion* in the presence of Cl^- ion to remove the intrinsic oxide layer and expose the anodic metal sites occurs as follows:

$$\left.\begin{array}{l} Al \rightarrow Al^{3+} + 3\bar{e}\ (\text{anodic site}) \\ 2H^+ + 2\bar{e} \rightarrow H_2\ (\text{cathodic site}) \end{array}\right\}. \tag{18}$$

When the product $(Al^{3+})(OH^-)_3$ reaches the saturation limit of aluminum hydroxide, the anodic reaction becomes

$$Al + 20H \rightarrow\ Al\,(OH)_3 + 3\bar{e}. \tag{19}$$

In the presence of dissolved oxygen, this reaction (Eq. 19) can be even more complex and form AlO (OH) products which, because of their semiconducting properties, can serve as more electropositive sites for copper deposition from solution (Obispo et al. 2000).

Fig. 9 Corroded aluminum can in acidic chloride solution during Cu cementation reaction (Eq. 17)

It might also be noted that aluminum can react aggressively with strong bases such as NaOH to produce hydrogen:

$$2\,\mathrm{Al} + 2\,\mathrm{NaOH} + 6\,\mathrm{H_2O} \rightarrow 2\mathrm{Na}\left[\mathrm{Al(OH)_4}\right] + 3\,\mathrm{H_2} \tag{20}$$

Aluminum in this context is referred to as an *amphoteric metal*. Beryllium, tin, and lead are also amphoteric metals which dissolve in concentrated NaOH solutions. Amphoteric metals and most metalloids can react with both acids and bases. They also form amphoteric oxides or hydroxides. For example, Al_2O_3 can react as follows:

$$\text{with acid:}\ \ \mathrm{Al_2O_3} + 3\,\mathrm{H_2O} + 6\,\mathrm{H_3O^+} \rightarrow 2[\mathrm{Al\,(CH_2O)_6}]^{3+} \tag{21}$$

$$\text{with base:}\ \ \mathrm{Al_2O_3} + 3\,\mathrm{H_2O} + 2\,\mathrm{OH} \rightarrow 2\left[\mathrm{Al(OH)_4}\right], \tag{22}$$

Similarly, ZnO can react as follows:

$$\left.\begin{array}{c}\text{with acid}:\ \mathrm{ZnO} + 2\mathrm{H^+} \rightarrow \mathrm{Zn^{2+}} + \mathrm{H_2O} \\ \text{with base}:\ \mathrm{ZnO} + \mathrm{H_2O} + 2\,\mathrm{OH} \rightarrow [\mathrm{Zn\,(OH_4)}]^{-2}\end{array}\right\}. \tag{23}$$

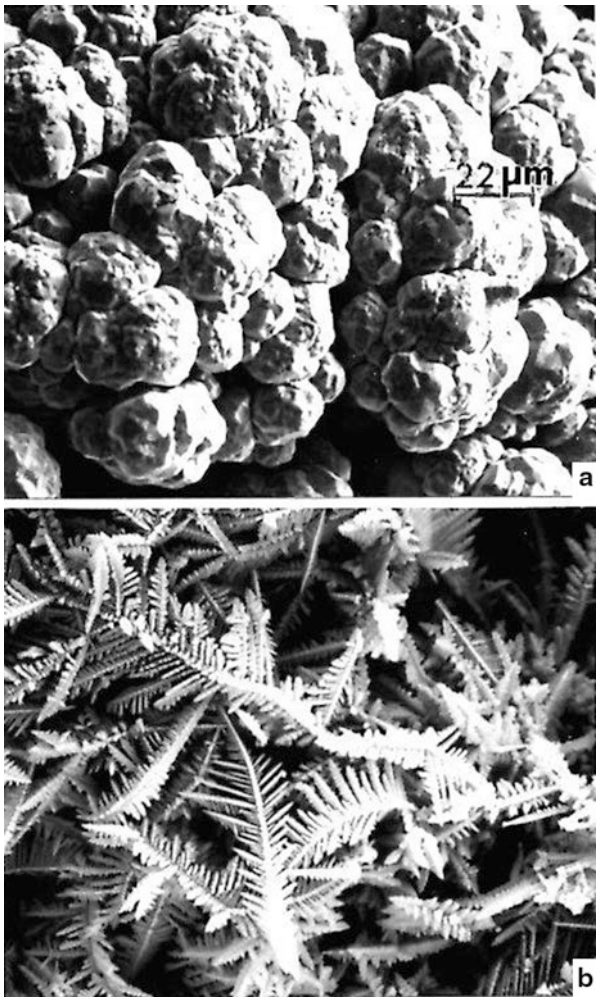


Fig. 10 SEM images of copper deposit structures on Fe (**a**) and Al (**b**). (**a**) Cu/Fe system deposit after 30 min. in 2 g/L Cu^{2+} solution at pH 2.2; 50 °C. (**b**) Cu/Al system deposit after 30 min. in 0.1 g/L Cu^{2+} solution containing 75 mg/L Cl^{-}; pH 3.2 at 50 °C (Adapted from Murr and Annamalai 1978)

Vapor, Liquid, and Solid - State Embrittlement of Materials

As illustrated for intergranular fracture in Ir by oxygen diffusion along the grain boundaries in ▶ Fig. 30 in the previous chapter "Fracture Modes and Mechanisms," along with similar examples of intergranular fracture in steels shown in ▶ Fig. 31 of chapter "Fracture Modes and Mechanisms," embrittlement by diffusing species such as hydrogen or other elements such as phosphorous, especially along grain boundaries where diffusivities are large (▶ Fig. 12 of chapter "Fracture Modes and Mechanisms"), can contribute to crack-assisted propagation by complex reactions. These promote stress-corrosion crack growth, or the creation of pressurized gas bubbles which magnify the crack tip stress concentration and crack propagation

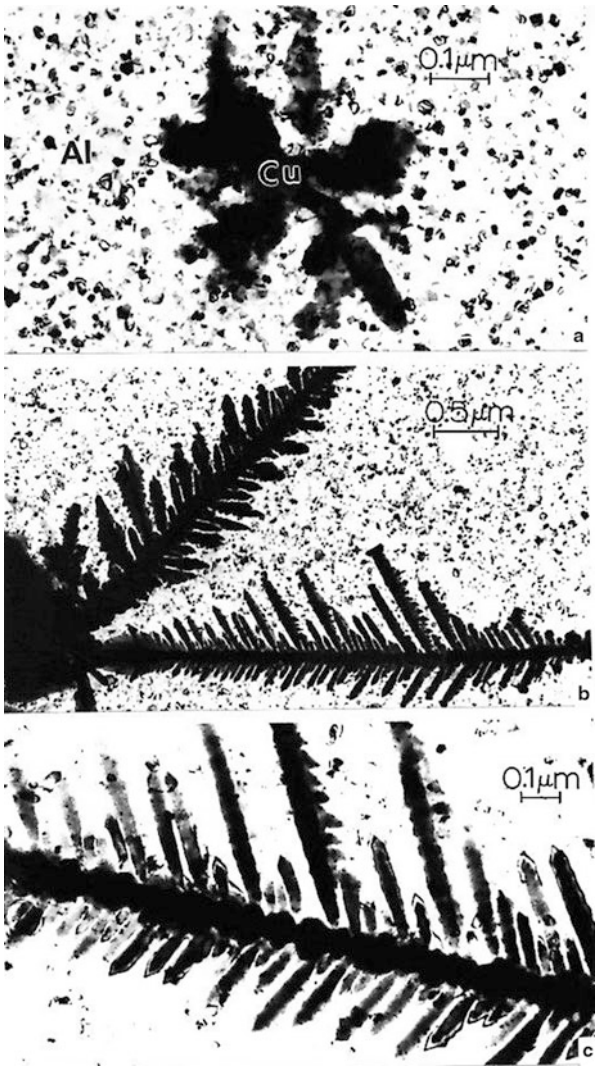


Fig. 11 TEM images showing Cu nucleation (deposition) and growth on Al thin film surface contacting $CuSO_4$ solution. (**a**) Microdendritic nucleus after 10 s contact at 30 °C. (**b**) Advanced microdendrite growth after 20 s exposure at 50 °C. (**c**) Shows a magnified view of dendrite arm structure in (**b**) (From Murr and Annamalai 1978)

assistance. In addition, hydrogen generated during corrosion of steel pipes and other underground storage facilities can migrate to surface cracks causing the same accelerated crack propagation phenomena. Hydrogen can also be introduced during processing of steels, aluminum, and Ti alloys in particular, especially at elevated temperatures.

In alloys such as stainless steels which are often sensitized by precipitation of carbides at the grain boundaries, diffusing hydrogen can react with the carbon to create methane (CH_4) in the grain boundary which is immobile, and which can build up enormous pressure in the boundary plane, causing catastrophic intergranular crack propagation. Similarly, diffusing hydrogen in Ti alloys or

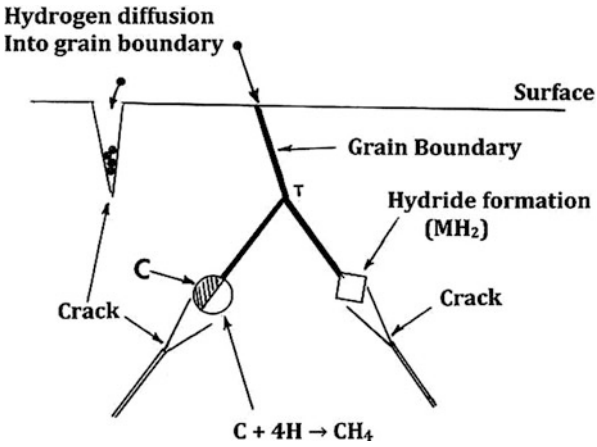


Fig. 12 Schematic showing examples of hydrogen embrittlement mechanisms

aluminum and aluminum alloys can react to form metal hybrids (MH_2) or more complex hybrids such as Ti-Fe H_2 which can also nucleate cracks either within an interface or in the matrix.

Figure 12 shows a general schematic illustrating some of the prominent hydrogen embrittlement mechanisms, while Fig. 13 illustrates two examples of intergranular failure by hydrogen embrittlement in high-strength steels. Figure 13a shows a zinc-plated carbon steel component where hydrogen can be introduced during the plating process. Zinc can also corrode in acidic environments releasing hydrogen as illustrated in substituting Zn for Fe in Eq. 13. Hydrogen created in these reactions can also diffuse to form blisters or internal bubbles which can develop considerable pressure, forming stress concentrations and crack nuclei. Depending upon the propensity of hydrogen bubble formation, cracking can be either transgranular or intergranular. It can be noted in Fig. 13b that there are remnants of ductile (dimple) fracture features on the fracture surface.

Figure 12 also illustrates the intersection of grain boundaries at triple points (T) where crack formation may be particularly prominent. Correspondingly, nanograin materials may also be especially prone to hydrogen-induced embrittlement as implicit on comparing Figs. 10 and 13 in the previous chapter "► Fracture Modes and Mechanisms."

It might also be mentioned that fission and fusion reactor environments generate significant amounts of both helium and hydrogen, with hydrogen arising primarily from fast neutron reactions. Both are known to exert deleterious effects on metals such as stainless steels through bubble formation and the creation of void cavities, thereby creating void swelling. Hydrogen has also been suggested to play a role in irradiation-assisted stress-corrosion cracking as well.

While the recognition of hydrogen embrittlement dates to around 1875, failures of metal structures by *liquid metal embrittlement* are much more recent. This is especially pertinent since advanced nuclear reactors, often referred to as liquid metal fast reactors, utilize liquid metal as the primary coolant. These are especially

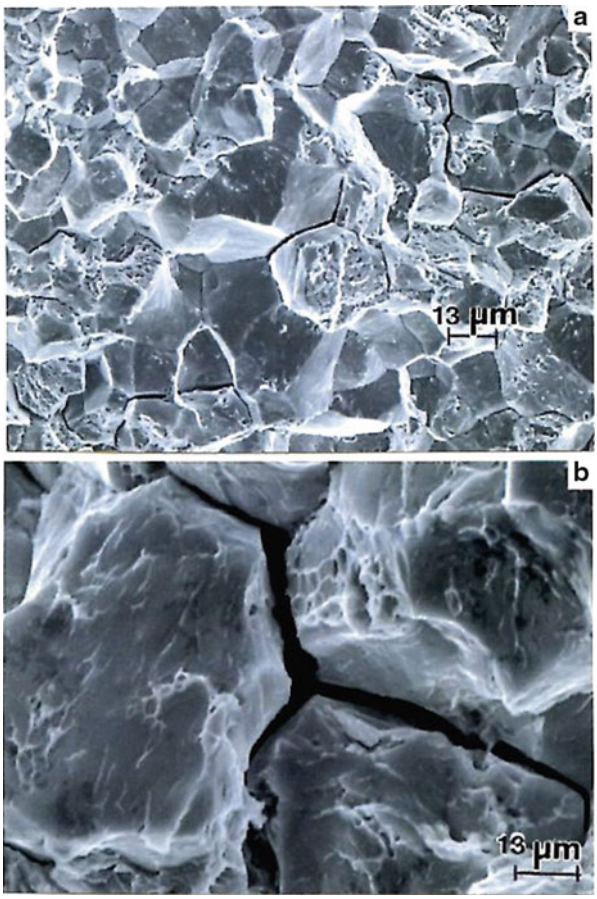


Fig. 13 SEM images showing hydrogen embrittlement in steels. (**a**) Zinc-plated carbon steel (Courtesy Eaton Aerospace Laboratories, Cleveland, OH). (**b**) High-strength steel (Courtesy Metallurgical Associates, Inc., Waukesha, WI)

prominent in naval propulsion systems in submarines. Prominent among such liquid metal coolants are Na, Hg, Pb, and Sn. These can embrittle metals and alloys, and judicious choices of reactor materials are essential.

The embrittlement of brass by Hg was one of the earliest systems studied (Huntington 1914), followed by Hg-embrittled aluminum, Bi- and Ga-embrittled Cu, and Zn-embrittled steel, including stainless steel (Nicholas and Old 1979). Figure 14 illustrates a novel example of liquid metal embrittlement where a galvanized zinc coating on a low carbon steel melted and penetrated the grain boundaries, causing catastrophic intergranular, smooth (brittle) fracture.

Grain boundary cracking as illustrated in ▶ Figs. 30, ▶ 31 (of chapter "Fracture Modes and Mechanisms"), 13, and 14 all require some applied stress to affect fracture. In addition, while it is implicit in Fig. 12 that stress concentration phenomena can play a significant role, surface (interfacial) adsorption also plays a significant and often corresponding role. Fundamentally, surface adsorption, especially implicit in oxygen adsorption in ▶ Fig. 30 of chapter "Fracture Modes

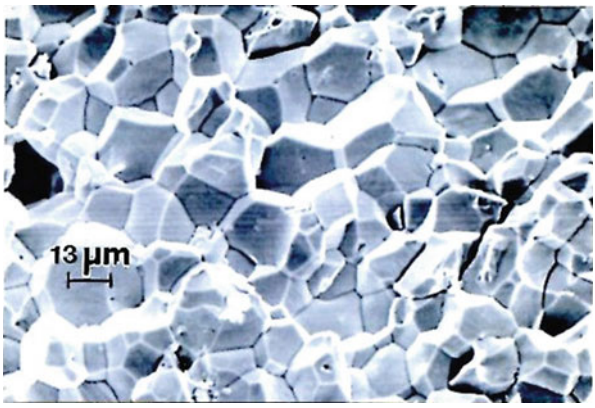


Fig. 14 Liquid zinc embrittlement of low carbon steel. SEM image showing intergranular fracture surface (Courtesy of Affiliated MRD, LLC, Edison, NJ)

and Mechanisms," significantly lowers the interfacial free energy. Shewman (1998) has noted that "equilibrium segregation of impurities at surfaces, and the concomitant reduction in surface energy, plays a central role in most, or at least many, types of grain boundary cracking." In retrospect, fracture phenomena are often complex phenomena.

References

Huntington AK (1914) Embrittlement of brass by mercury. J Inst Metals 11:108–112

Murr LE (1980) Theory and practice of copper sulfide leaching in dumps and in-situ. Miner Sci Eng 12(3):121–189

Murr LE, Annamalai V (1978) An electron microscopic study of nucleation and growth in electrochemical displacement reactions: a comparison of the Cu/Fe and Cu/Al cementation systems. Metall Trans 9B:515–525

Nicholas MG, Old CF (1979) Review: liquid metal embrittlement. J Mater Sci 14:1–18

Obispo HM, Murr LE, Arrowood RM, Trillo EA (2000) Copper deposition during the corrosion of aluminum alloy 2024 in sodium chloride solutions. J Mater Sci 35:3479–3495

Shewmon PG (1998) Grain boundary cracking. Metall Mater Trans B 29B:509–518

Materials in Extreme Environments

Contents

Abstract

Extreme environments associated with turbine operation in power generation and turbines and turbine components as well as systems and components of nuclear reactors are discussed in the context of extremes in temperature, stress and stress variations, radiation, extreme chemical environments, and their complex interactions. Other energy production-related environments in oil and gas production and general chemical and high-pressure environments are presented. Special attention is devoted to space environments, especially low Earth orbit (LEO) where micrometeoroid and space debris impacts, atomic oxygen effects, radiation and thermal cycling, and related and interactive degradation features pose concerns and important design strategies.

Introduction

Extreme environments can involve temperature and heat flux extremes, extreme chemical environments, large ionizing and atomic particle or radiation fluxes, large and complex stress variations, electric and magnetic fields and large electromagnetic pulses, and especially combinations or other complex interactions of these extremes. In addition, gas, fluid, and solid interactions in combination with harsh chemical components and high pressures can cause combinations of corrosion, erosion, and catastrophic rupture of materials and materials systems. Such environments can exist in nuclear reactor systems, oil drilling, petrochemical systems and material components, turbine systems and components, and space, including low

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_55

Earth orbit (nominally ~ 180–650 km from the Earth's surface), and outer space. The design of materials to function in extreme environments is often predicated upon observations of their performance in such environments or in simulated environments where the performance effects of many extreme parameters can be selectively observed.

► Figure 38 in chapter "Fracture Modes and Mechanisms" serves to illustrate the extreme environments posed by turbine systems operating in very high-temperature combustion-related environments where combinations of erosion, high temperature, corrosion, and stress can act simultaneously on blade structures and other critical components to cause catastrophic failures having enormous impact or significant consequences. ► Figures 8 and ► 12, etc., in chapter "Examples of Directional Crystal Structures: Gas Turbine Component Applications in Superalloys" also discuss turbine blade and component development and degradation as well as implications for future materials requirements and optimization as the demands of power systems and especially their operational temperatures increase.

Nuclear power systems, especially fission reactors such as pressurized water reactors (PWRs) and boiling water reactors (BWRs), pose particular challenges for materials development and performance enhancement especially in light of complex temperature, corrosive/erosive effects, stress, and radiation interactions. As illustrated in ► Fig. 9 in chapter "Volume Defects: 3D Imperfections in Crystals," helium and hydrogen (Garner and Greenwood 1998) are formed in reactor stainless steels, as well as other metals and alloys, by neutron-induced transmutation. Hydrogen can also be produced by other mechanisms including corrosion, recoil injection of protons after neutron-water collisions, equilibrium dissociation arising from hydrogen overpressures in PWRs, and radiolytic decomposition of water in light water reactors (LWRs). Both helium and hydrogen can stabilize vacancy clusters forming bubbles and void cavities which produce void swelling. This phenomena combined with irradiation-assisted stress corrosion cracking (SCC) can have catastrophic effects in nuclear reactors and pose considerable challenges for future fusion reactor environments (Jenssen and Ljungberg 1995; Gordon 2013; Caro et al. 2013).

While these and other power and energy production systems and environments pose considerable materials challenges (Hemley et al. 2009; Samaras et al. 2011), a particular challenge for materials performance and development exists in space-related activities. These are poised to involve considerable private, economic ventures in contrast to government (NASA)-driven space ventures which have highlighted the past more than half-century of spacecraft and space systems development, including the international space station.

Materials Effects in Low Earth Orbit (LEO)

The launch of the Soviet satellite Sputnik I in 1957 started the so-called space race with the USA, which created the Project Mercury program led by the National Aeronautics and Space Administration (NASA) with the goal of putting a human in

orbit around the Earth first. However, the Soviets put Yuri Gagarin into orbit in April 1961, followed by John Glenn of the USA in May 1961. From 1962 to 1966, the Project Mercury program was followed by the two-man Project Gemini. Correspondingly, the Apollo program of manned space flight, the third NASA human spaceflight program, ran from 1961 to 1972 with the goal of manned lunar landing using Saturn family rockets as launch vehicles utilizing liquid oxygen and hydrogen fuels. Unmanned Apollo flights to test launch and module performance, including heat shield reentry performance, were conducted between 1966 and 1968, with the first manned (three astronauts, Shivra, Cunningham, and Eisele) Earth-orbited mission in October 1958 (Apollo 7). The first manned landing on the Moon by Neil Armstrong, Buzz Aldrin, and Michael Collins occurred in July 1969 (Apollo 11). The sixth and final Moon landing occurred in December 1972 (Apollo 17).

The NASA Space Shuttle Program or Space Transportation System (STS) began in 1972. The winged space shuttle orbiter was launched vertically and carried from 2 to 8 astronauts and up to 22,700 Kg payloads. It was initially conceived as a means to build a space station in low Earth orbit (LEO) and serve as a supply vehicle. The first US space station, Skylab, orbited the Earth from 1973 to 1979. It was launched unmanned by a Saturn V rocket. Three manned missions to the station were conducted between 1973 and 1974 using the Apollo program module with a three-astronaut crew. Skylab reentered Earth's atmosphere and disintegrated in 1979. Following Soviet and later Russian Salyst, Almaz, and Mir space stations, the International Space Station (ISS), a modular structure, launched its first component in 1998. It has been continually occupied since 2000 by astronauts and cosmonauts from 15 different nations and serviced by a variety of government and private (commercial) spacecraft. The ISS maintains an orbital altitude of between 330 and 435 km (270 mi) using module engines or visiting spacecraft. Currently funded to 2020, the ISS is touted to be the most expensive single item ever constructed, at a cost of ~ $160 billion dollars, although the total shuttle program cost was estimated in 2012 to be nearly $200 billion dollars.

Two critical materials failures caused shuttle accidents which killed all astronauts aboard: the Challenger (STS-STL) at liftoff in 1986 and the Columbia (STS-107) at reentry in 2003. The liftoff tragedy in 1986 occurred because of a thermally embrittled liquid oxygen fuel tank O-ring, while the 2003 reentry tragedy occurred because a piece of iced foam shed from the external fuel tank struck the leading edge of the orbiter's left wing, puncturing one of the reinforced carbon-carbon reentry protection panels. On reentry, hot gas penetrated this unprotected wing section causing internal damage which led to loss of vehicle control and eventual complete disintegration only ~ 16 min before landing, creating a debris field stretching over most of Texas and part of Louisiana in the USA. The final shuttle space launch of the Shuttle Atlantis occurred on July 8 2011.

The first attempts to explore LEO occurred around 1946, after World War II, when captured German V-2 rocket and related technology were employed by the USA, to examine cosmic radiation effects and photograph the Earth from space. The Soviets followed in 1947. There is some evidence that the Germans

considered the prospects for the extreme environment in LEO in designing the V-2 rocket as early as 1942. In 1947 Fred Whipple (Whipple 1947) invented a type of hypervelocity impact shield (Whipple shield or Whipple bumper) for protecting spacecraft from micrometeorite particles and orbital debris particle impact at velocities $\geq$5 km/s. These shields consisted of spaced layers which would progressively fragment and collect impacting particles at hypervelocity. Ultimately, space suit fabric was also constructed in spaced layers to protect astronaut activities outside spacecraft.

In the initial stages of the US space program, the primary concerns involved light-weight materials and structures designed to withstand micrometeorite impact, sublimation and outgassing in LEO, thermal and radiation effects, and atomic oxygen (O) which was known to be a particular hazard beyond the Earth's ozone (O_3) layer. Intense programs involved materials selection, configurations, analysis, and performance testing and evaluation which characterized the initial Mercury spacecraft design which had an outer skin made of Rene 41, a nickel-based superalloy (see ▶ Table 1, chapter "Examples of Directional Crystal Structures: Gas Turbine Component Applications in Superalloys") because of anticipated high temperatures especially during reentry. The heat shield was composed of an aluminum honeycomb. Because of the payload requirements, weight-saving aluminum alloys were employed in Apollo program modules which required special attention to developing an effective thermal protection system which became especially effective in the subsequent STS program shuttle craft fabrication. Seven different thermal protection materials systems covered essentially the entire shuttle orbiter surface: reinforced carbon-carbon in the shuttle nose (where reentry temperatures exceeded ~ 1,260 °C), reusable surface insulation tiles made of silica-ceramic where temperatures were <650 °C, low-temperature tiles in the same range of temperatures as blankets, and two other low-temperature fibrous insulation materials. The silica-ceramic tiles covered most of the surface of the shuttle as illustrated in Fig. 1.

Although earlier studies had contributed to successful design and development of unmanned and manned modules (Parker 1963; Purser et al. 1964; Osgood 1966), the NASA Langley Research Center proposed more extensive experiments especially of a wide range of materials effects in LEO in light of proposals for Skylab and other space station ventures. Led by NASA scientist William H. Kinard, the Long Duration Exposure Facility (LDEF) concept was proposed in 1970 and approved for development in 1974. Initially focused on constructing a Meteoroid and Exposure Module, LDEF became a hollow, 12-sided cylinder about the size of a bus (9.2 m in length and 4.6 m diameter). Each of the 12 sides contained rows of trays attached by clamps, with additional trays attached to each end of the cylinder. The 12 rows of trays represented 57 different experiments and more than 10,000 different test specimens of materials of all types: metals and alloys, ceramics, polymers, composites (especially fiber-reinforced polymers), semiconductors of various types and configurations, including circuit modules, etc. Figure 2 shows a view of the LDEF spacecraft after retrieval and returning to the NASA Kennedy Space Center for initial inspection, while Fig. 3 shows a few examples of the

Fig. 1 NASA space shuttle orbiter Atlantis at launch showing solid rocket/liquid oxygen/hydrogen rocket configuration (NASA photograph)

materials trays prior to deployment. The LDEF spacecraft was deployed in LEO by the NASA Challenger shuttle orbiter on April 7 1984 (Fig. 4). While it was intended to be retrieved after 1 year, schedule changes and the Challenger tragedy in 1986, followed by shuttle flight suspensions, allowed LDEF to remain in LEO for 69 months. It was retrieved by the shuttle orbiter Columbia in January 12, 1990, after circling the Earth nearly 34,000 times (Fig. 2). LDEF's initial orbit at a height of 433 km (260 mi) was nearly circular, but it was decaying rapidly (~2 km/day) at the time of its retrieval to an irregular orbit as low as 300 km (~180 mi).

Several key issues, in addition to analyzing micrometeoroid impact, were part of LDEF's test and analysis goals. These included studies of ionizing radiation, especially effects on electronic components; deleterious effects of atomic oxygen, especially on polymeric composites; and particulate contamination and debris production.

After more than 3 months of careful photographing of the entire 130 m^2 area of LDEF, and subsequent characterization of micrometeoroid and related hypervelocity particulate impact craters, it was estimated that there were more than a billion craters larger than ~ 10 nm in diameter (See et al. 1990). A few examples are illustrated in Figs. 5 and 6.

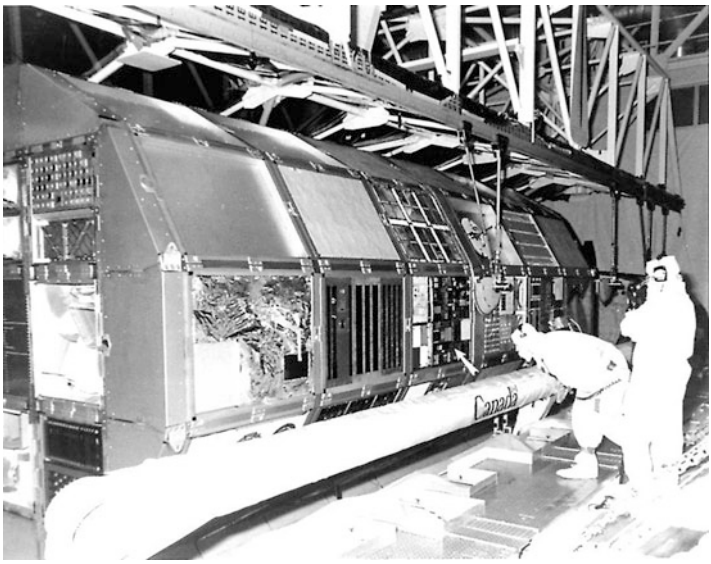


Fig. 2 LDEF inspection after retrieval from orbit in January 1990 at John F. Kennedy Space Center. The space-facing end is at the *left*. *Arrow* in front of technician indicates tray D in row 9 of leading edge of spacecraft in space (NASA photograph)

Figure 5 shows two 303 stainless steel bolts arranged side-by-side showing impact craters. These bolts fastened anodized aluminum alloy tray clamps: each of 12 rows of 6 trays and 14 trays in the space and Earth-facing ends of LDEF. Figure 6a, b shows craters in the anodized aluminum alloy (Al-6061) surface of LDEF. While most of the exposed surface of LDEF was anodized Al-6061, some sections contained unanodized plates of the alloy. In addition 17 of the peripheral trays were covered with thermal blankets consisting of an outer Teflon layer backed with sputter-deposited silver/Inconel, which in turn was backed with primer and a black, conductive urethane paint. Hypervelocity impact on these multilayer blankets often penetrated the complete composite thickness and produced complex delamination (Liu et al. 1993) as illustrated for a hypervelocity impact crater on a high-reflectance mirror containing a multilayer ZnS/Al_2O_3 coating as shown in Fig. 6c.

Examination of a wide array of impact craters on various LDEF components and surfaces indicated that apparently many were "self-inflicted debris" impacts caused by primary impacts producing debris by rim ejection, etc., which re-impacted the spacecraft, along with other orbital debris. Other deep space particle impacts were also detected, and the mean impact velocity was estimated to be ~ 16–17 km/s (Murr and Kinard 1993). Space debris of various kinds and origins, including impact and erosion products, are believed to compose roughly half of the particulate loadings in LEO. In fact, more than a trillion contaminating particulates of various kinds were estimated to be on the surface of the retrieved LDEF spacecraft (Murr and Kinard 1993).

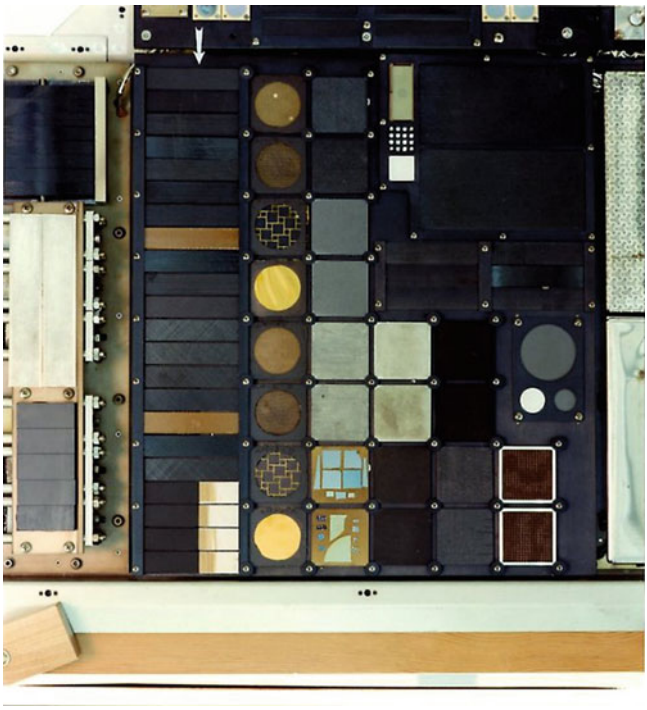


Fig. 3 LDEF experiment (materials) tray (D09) section prior to deployment in area shown at *arrow* in Fig. 2. *Arrow* at *upper left* shows a row of polymer-matrix composite materials for evaluation in LEO environment

Earlier shuttle experiments had demonstrated deleterious materials effects of atomic oxygen (Leger and Visentine 1986), and LDEF contained a wide range of materials to further evaluate atomic oxygen effects. These included polymers and polymer-matrix composites of various kinds which had been shown to be especially affected: graphite polyimide and polysulfone, carbon phenolic, quartz fiber phenolics, and common polymers such as polymethyl methacrylate (PMMA). Many of these polymer composites were located in the tray shown in Figs. 2 and 3. This tray, shown shaded (D09) in Fig. 7, was on the leading edge of the spacecraft where the fluence of atomic oxygen (O) was greatest (~10^{22} O atoms/cm^2) in contrast to the trailing edge where the flux of atomic oxygen was ~ 10^4 O atoms/cm^2.

Figure 8 illustrates some typical examples of severe atomic oxygen erosion damage on the leading edge of LDEF (Fig. 7). In Fig. 8a the erosion of resin in a white thermal-control paint on a tray clamp exposed mineral components: talc platelets at T and TiO_2 white pigment spherical particles and aggregates. Figure 8b shows atomic oxygen erosion of a graphite-epoxy composite which was typical of most polymers and polymer-matrix composites observed especially

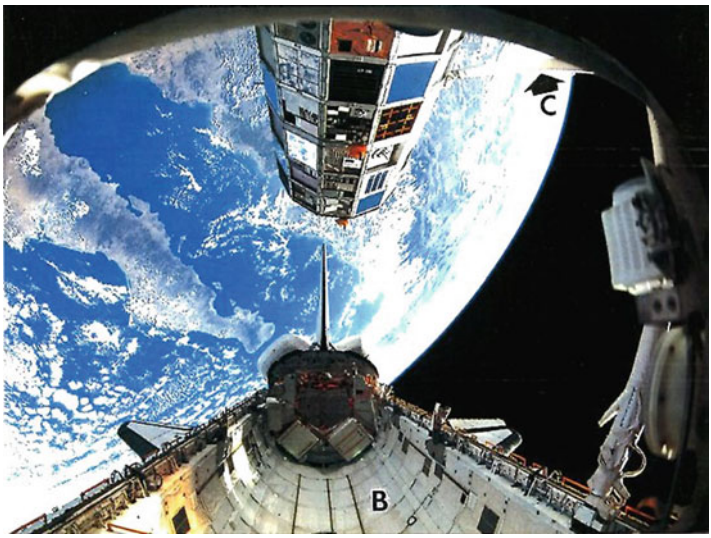


Fig. 4 LDEF after lifting from the shuttle orbiter Challenger's payload bay (*B*) on April 7, 1984, using the Shuttle Remote Manipulator System (*SRMS*) or Canadarm (*C*), a mechanical arm to maneuver payloads from the orbiter payload bay to the deployment position in orbit (NASA photograph looking toward tail section of shuttle)

on the spacecraft leading edge (Fig. 7). Correspondingly, Fig. 8c–f shows the erosion effects of atomic oxygen on red PMMA bicycle reflectors mounted on the trailing (Fig. 8c, e) edge in contrast to the leading edge (Fig. 8d, f).

The atomic oxygen erosion is an energetic process involving the collision of mono-oxygen atoms at orbital velocity (~8 km/s) corresponding to ~ 5 eV collision energy. This can be accelerated and further catalyzed by UV radiation and thermal fluctuations or cycles. Atomic oxygen fluences are more severe near the upper ozone layer corresponding to lower LEO where LDEF spent its final months in orbit. Many thousands of geosynchronous and non-geosynchronous satellites in LEO have been subject to many of these effects.

Since LDEF, several series of materials test and evaluation experiments have been mounted externally on the International Space Station (ISS; Fig. 9) to continue investigating long-term exposure to space environment. Roughly 1,500 materials samples are being investigated in what has become known as the Materials International Space Station Experiment (MISSE) series. Participants include NASA facilities (Langley Research Center, Johnson Space Center, Marshall Space Flight Center, and Glenn Research Center), the Materials Laboratory at the Air Force Research Laboratory, Boeing Phantom Works, and Infosight Corporation.

In 2013, former NASA astronaut Danny Olivas and colleagues at the University of Texas at El Paso founded the Center for the Advancement of Space Safety and Mission Assurance Research (CASSMAR) to critically examine an array of collected debris from the Columbia (STS-107) reentry disaster in 2003 (Fig. 10).

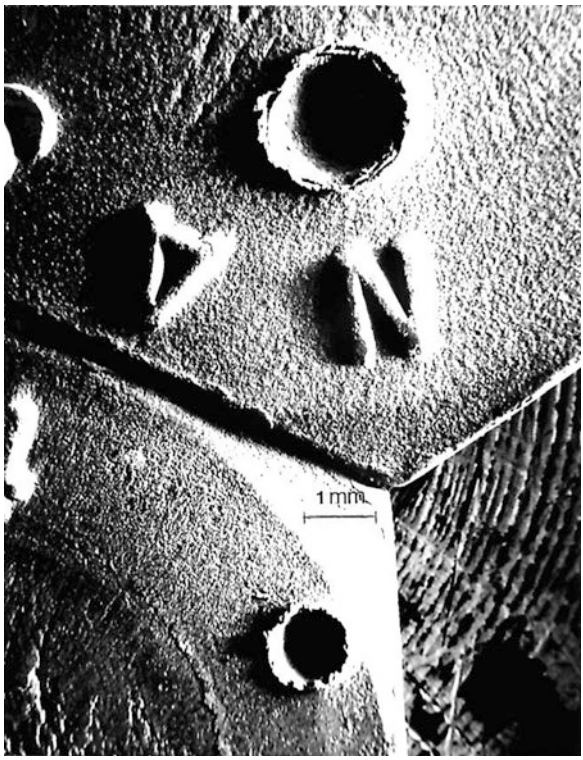


Fig. 5 Impact craters in 303 stainless steel hex bolts used to fasten trays on LDEF. Bolt heads are placed side by side for imaging

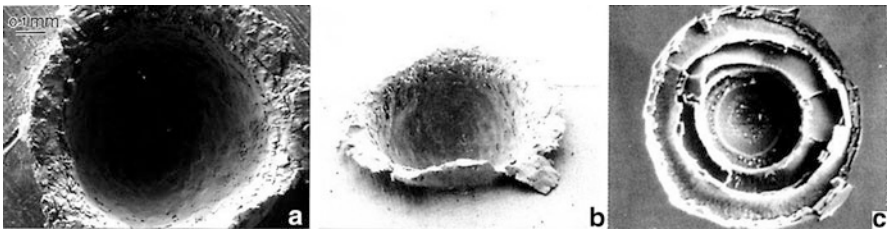


Fig. 6 Examples of hypervelocity projectile (particle) impact craters on LEDF components. (**a**) Anodized aluminum alloy surface area. (**b**) Unanodized aluminum alloy surface areas. (**c**) Impact and penetration into multilayer ZnS/Al_2O_3 coating on high-reflectance mirror surface (Adapted from Murr and Kinard (1993))

The main thrust of this effort is to examine unusual materials performance, degradation, and failure in the context of prevention and optimization for future space vehicle and structure design, especially in the commercial (private) space exploration arena. This is especially important for reentry materials and materials configurations as well as the consideration of materials performance beyond LEO and especially long durations in space.

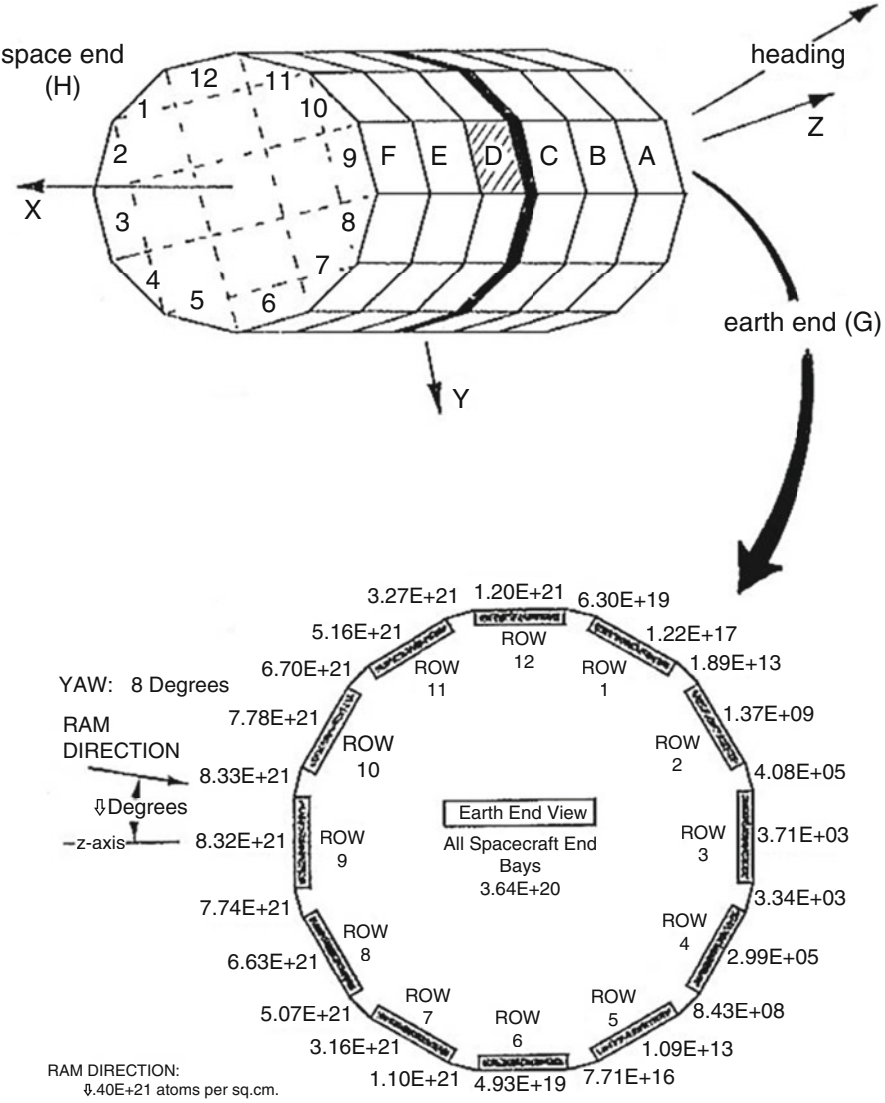


Fig. 7 Schematic view of LDEF spacecraft oriented as shown in Fig. 2 showing orbital configuration and designation of trays along rows (*top*). *Bottom* shows the atomic oxygen fluences (Courtesy of NASA)

Previous examination of *Columbia* space shuttle debris (Figs. 10 and 11) by Mayeaux et al. (2004) in the leading edge of the breached left wing indicated temperatures inside structural subsystems exceeded 1,760 °C which resulted in complex melt deposits on many components. One of the more unique materials issues involved *liquation* of aluminum alloy debris as a result of this high

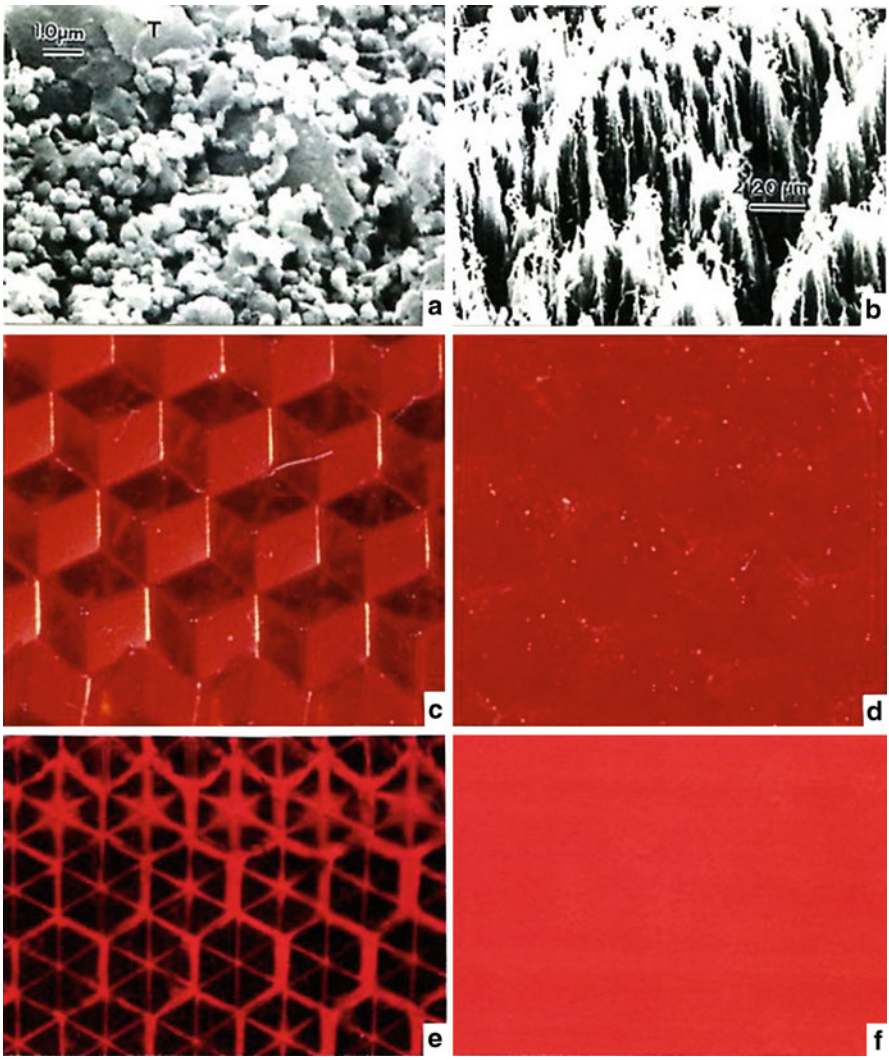


Fig. 8 Atomic oxygen damage/degradation of LDEF materials. (**a**) Paint erosion on tray clamp. (**b**) Graphite-epoxy composite erosion. (**c**) Trailing edge PMMA reflector (reflected light). (**d**) Leading edge PMMA reflector showing erosion damage. (**e**) Trailing edge (and initial) PMMA reflector (transmitted light). (**f**) Leading edge PMMA reflector in (**e**) (transmitted light) showing erosion damage causing no light transmission (Adapted from Murr and Kinard (1993))

temperature (Lam et al. 2006), which resulted in brittle failure characteristics. Liquation or liquation-induced grain boundary segregation and embrittlement is similar in some respects to liquid metal embrittlement described previously. It results by the selective melting of low-melting-point precipitates (often eutectics) within the grain boundary and may be due fundamentally to the nanodimensional

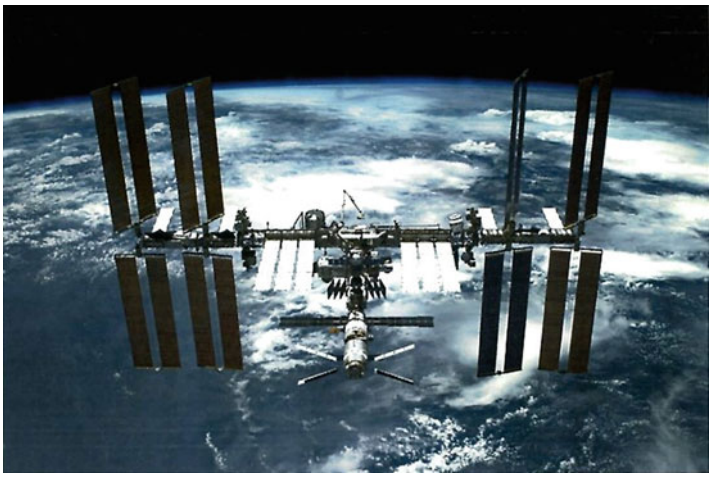


Fig. 9 International Space Station (ISS) in orbit. The ISS makes 15.5 orbits per day. The orbital altitude ranges from ~330 to 445 km above the Earth. The width shown is ~108 m. Crew of 6 occupied the station in 2013 (NASA photograph from Wikimedia Commons)

Fig. 10 *Columbia* debris reconstruction grid in the NASA Reusable Launch Vehicle Hangar. Roughly 40 % of the shuttle weight was recovered in debris (NASA photograph)

suppression of the melting point as illustrated in ▶ Fig. 7 in chapter "3D Printing: Printed Electronics," as well as the melting of nano-thin ductile-dimple structures in high-strain-rate fracture illustrated in ▶ Fig. 37 of chapter "Fracture Modes and Mechanisms" above. This phenomenon has been observed in welded aluminum alloys as well as thermal treatments of other alloy systems (Huang and Kou 2002; Zhong et al. 2005).

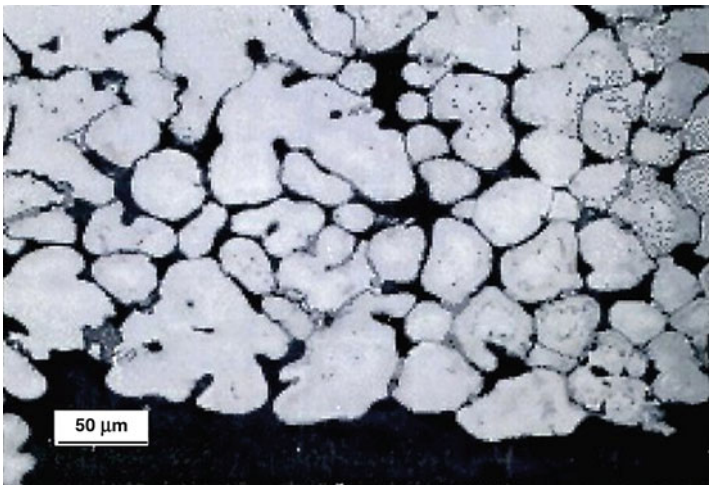


Fig. 11 Grain rounding and extreme grain boundary liquation at the fracture surface in a cast aluminum alloy structural member containing grain boundary precipitates (Fe, Mn, Ni, Ag composition). From space shuttle *Columbia* debris (Fig. 10) (Adapted from Lam et al. (2006))

References

Caro M, Woloshon K, Rubio F, Malloy SA, Hosemann P (2013) Heavy liquid metal corrosion of structural materials in advanced nuclear systems. JOM 65(8):1057–1066

Garner FA, Greenwood LR (1998) Neutron irradiation effects on fusion or spallation structural materials: some recent insights related to neutron spectra. Radiat Eff Defect Solid 144:251–283

Gordon BM (2013) Corrosion and corrosion control in light water reactors. JOM 65(8):1043–1056

Hemley RJ, Crabtree GW, Buchanan MV (2009) Energy challenges for materials in extreme environments. Phys Today 62:32–37

Huang C, Kou S (2002) Liquation mechanisms in multicomponent aluminum alloys during welding. Welding J 2002:211-S–222-S

Jenssen A, Ljungberg LG (1995) Final report: Irradiation assisted stress corrosion cracking studies, ABB Atom, SDA-95-1262, Washington, DC

Lam I, Ganot G, Brundage D, Velcheck L, Deacon RM, Marder AF (2006) Failure analysis of aluminum structural members in the space shuttle *Columbia*. J Fail Anal Prevent 6(1):32–38

Leger LJ, Visentine JT (1986) A consideration of atomic oxygen interactions with the space station. J Spacecraft & Rockets 23:50–56

Liu J, Niou C-S, Murr LE, Wang PW (1993) Hypervelocity particle penetration in multi-layered thermal blankets. Scri Metall Mater 28:377–382

Mayeaux BM, Collins TE, Jerman GA, McDanels SJ, Piascik RS, Russell RW, Shah SR (2004) Materials analysis: a key to unlocking the mystery of the *Columbia* tragedy. JOM 56(2):20–30

Murr LE, Kinard WH (1993) Effects of low Earth orbit. Am Sci 81:152–165

Osgood CC (1966) Spacecraft structures. Prentice-Hall, Englewood Cliffs

Parker ER (1963) Materials for missiles and spacecraft. McGraw-Hill Book, New York

Purser PE, Faget MA, Smith MF (eds) (1964) Manned spacecraft: engineering design and operation. Fairchild Publications, New York

Samaras M, Derlet PM, Swygenhoven H (2011) Materials challenges for generation IV nuclear energy systems. Nuclear Mater 10:608–613

See T, Allbrooks M, Atkinson D, Simon C, Zolensky M (1990) A preliminary report – meteoroid and debris impact features documented on the long duration exposure facility. NASA-JSC publication no 24608, Aug 1990

Whipple F (1947) Meteorites and space travel. Astronomical J 52:131–137

Zhong M, Sun H, Liu WJ, Zhu X, He J (2005) Boundary liquation and interface cracking characterization in laser deposition of Inconel 738 on directionally solidified Ni-based superalloy. Scripta Mater 53:159–164

Failure of Integrated Circuits

Contents

Abstract

Although there are many complex, integrated, and interactive systems such as automobiles, aircraft, electrical generation, and distribution systems, integrated circuits and especially very-large-scale integrated (VLSI) and ultra-large-scale integrated (ULSI) device systems demonstrate a length-scale system, in contrast, which is unique because of its continuous evolution. This evolution, driven for roughly four decades by Moore's law, involves feature sizes driven to the low-nanometer-size regime and transistor densities exceeding billions/cm^2. Electromigration is discussed in the context of these phenomena along with related circuit failure phenomena as these relate to product reliability.

Introduction

In 2013, integrated circuits (ICs) or ultra-large-scale integrated (ULSI) circuits contained ~5 billion transistors in a chip area <400 mm^2, using feature sizes of ~20 nm. In such dense arrays, which include resistive, capacitive, and inductive components, interconnects, and other features, the propensity for failure becomes daunting. As shown in the overview in Fig. 1, Murphy's law is perhaps as important as Ohm's law and especially the materials corollary of Murphy's law: *If any material can fail, it will*. As with other material degradation and failure discussed and illustrated in chapters "▶ Fracture Modes and Mechanisms," "▶ Chemical and Electrochemical Degradation and Failure of Materials," and

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_56

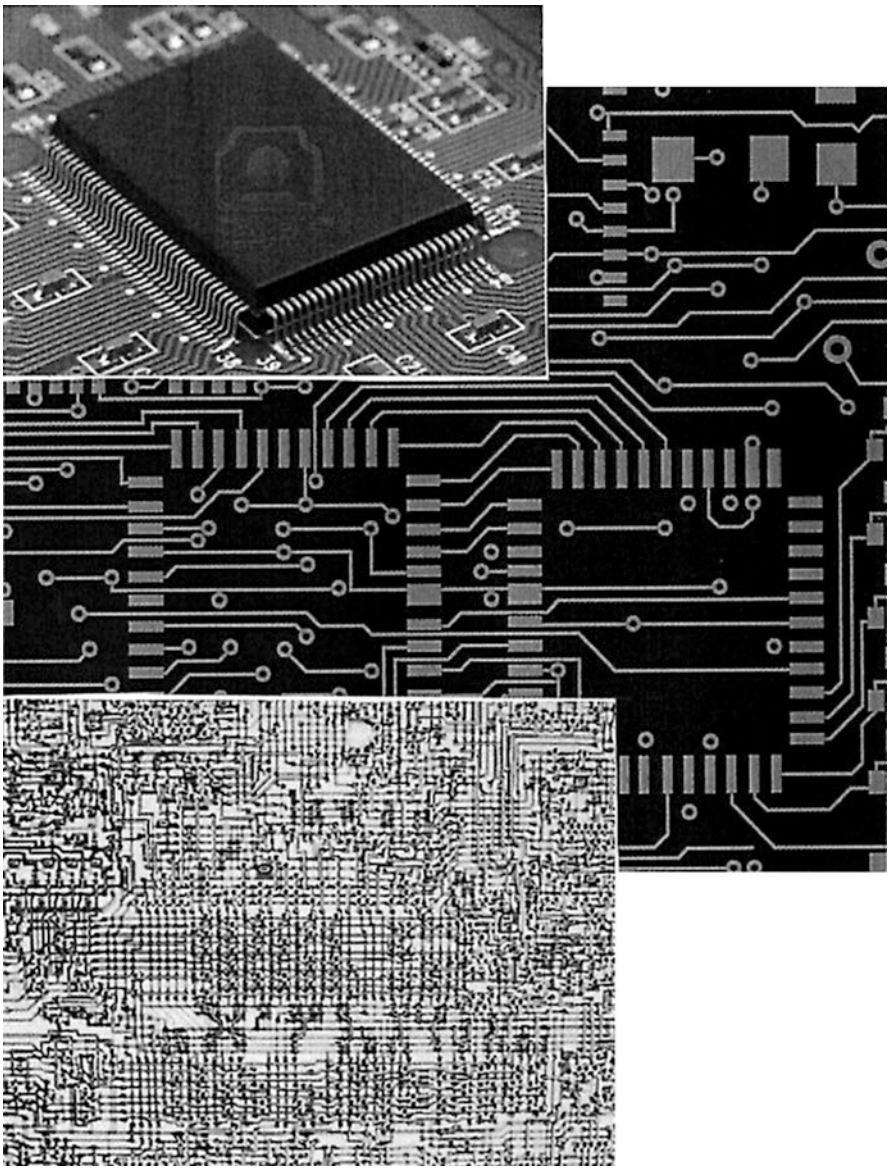


Fig. 1 Integrated circuit collage showing packaged die (*upper left*), exterior interconnects (runs), and chip architecture/components (*lower left*)

"► Materials in Extreme Environments" above, failure modes include a variety of stress modes, including thermal cycling and related thermal stresses, corrosion, and chemical and electrochemical reactions, including moisture and humidity-related phenomena, bond and interconnect failure, intermetallic formation and brittle fracture, ohmic contact degradation, and electromigration and related diffusional,

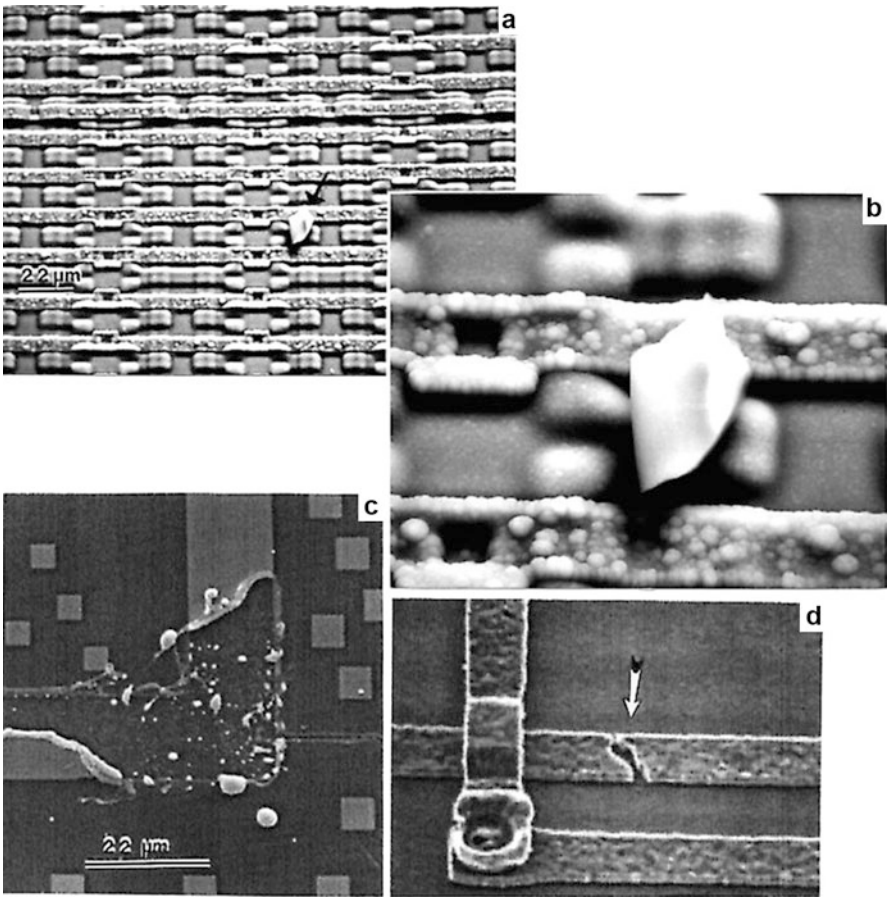


Fig. 2 Common IC degradation and failure examples. (**a**) and (**b**) show physical defect (impurity particle on interconnect area). (**c**) Interconnect melt due to current crowding. (**d**) Interconnect open line (possibly due to electromigration). Magnification of (**d**) is same as (**c**). (**b**) is a magnified view of (**a**) at *arrow*. Metallization interconnects shown are aluminum. (**d**) is courtesy of Daniel Payne

electronic, and ionic transport phenomena. These can lead to whisker formation and component shorting (short circuits) or void formation causing resistive heating, melting, or vaporization at reduced metallization cross sections. Electrostatic discharge (ESD) and oxide layer (insulator) breakdown, etc. can cause catastrophic damage at high discharge voltages. Many electronic parts failures are also packaging related, where moisture-related corrosion, thermal expansion, and cyclic thermal and mechanical stresses can cause thermal or thermomechanical fatigue of chip (die) sections. Radiation damage to packaging materials and the die components can also pose particular concerns in extreme environments such as low Earth orbit (LEO) (chapter "► Materials in Extreme Environments") or space environments beyond LEO.

Figure 2a, b illustrates physical defects such as a process remnant or rare dust particle (many die fabrication facilities require class 1 clean room conditions

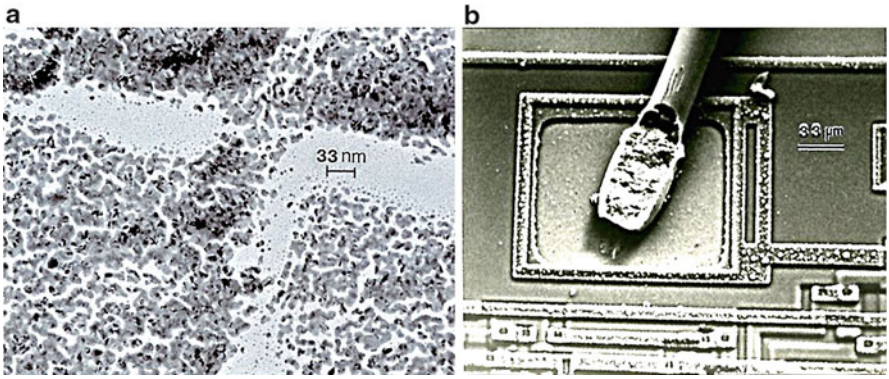


Fig. 3 (**a**) Gold deposition (vapor deposition) showing island structure and limited contact. (**b**) Al contact wire on Al pad in IC

($\sim$1 particle/m^3)). Figure 2c, d illustrates interconnect metal melting and open-circuit gap, respectively. Metallization using vapor deposition through photomasks requires parameter adjustments in substrate temperature and deposition rates since the depositing atoms often agglomerate rather than spread over the surface area, and this affects contact and conductivity. This feature is represented generally in the vapor deposition of gold as shown in Fig. 3a where gold islands $<$10 nm thick create complex, limited contact producing limited conductivity. This becomes a particularly important issue as feature size is reduced. Metallization cross-sectional areas in 2013 were roughly 10^3 nm^2 and estimated to decrease to half that in 5 years (2018). This will double the current density. Common interconnect metallization has been Al and W, which in very thin film interconnect structures has a much better conductivity than bulk W. Copper interconnects are also becoming more popular in ULSI die fabrication as well. Aluminum or gold wire is used to make the connections to integrated circuit front surfaces as illustrated in Fig. 3b. The formation of Au-Al intermetallic compounds sometimes poses concerns when gold wires are used, and aluminum wire bonding is used to obtain Al/Al interconnections, since the pads are usually aluminum. These are located around the perimeter or edges of the chip as illustrated in Fig. 3b.

Electromigration

Current flowing through conducting interconnects as shown in Fig. 2 produces two forces: a weak electrostatic force caused by the electric field strength in the metallic conductor and a more prominent momentum transfer force between the conduction electrons and the conductor atoms, the electrons pushing the atoms in what is often referred to as an *electron wind*. This electromigration of conductor atoms causes voids to form in one portion of the metallization cross section, while migrating atoms at the other portion of the metallization causes atom extrusion or hillock

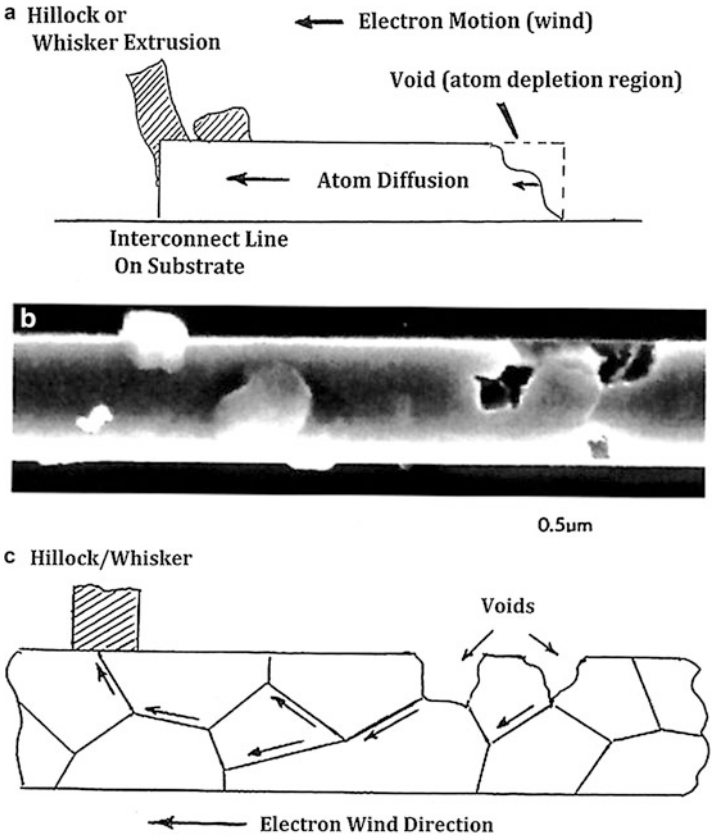


Fig. 4 Electromigration. (**a**) Schematic view of atom diffusion through metal interconnect line on IC substrate being "pushed" by the electron "wind." Loss of material at *right* forms void region. Transport of material to *left* forms hillock or extruded material. (**b**) SEM image of electromigration features in an Al interconnect line section. (**c**) Schematic of polycrystalline interconnect line microstructure in (**b**) showing material (Al) transport by grain boundaries oriented in the electron wind direction. Diffusional transport can also include dislocations or the lattice as shown in the chapter "▶ Fracture Modes and Mechanisms," Fig. 12

formation which can be directed into the formation of a whisker. This whisker extension can cause an unwanted connection or short circuit with another interconnect or chip component, while the void-forming area can cause an open-circuit condition shown in Fig. 2c. Figure 4a shows this electromigration phenomenon schematically while Fig. 4b illustrates actual void and hillock formation in a conducting interconnect segment by electromigration (EM).

The crystal structure or microstructure of the metallization has a significant effect on electromigration since diffusion plays a primary role in atom displacement in the electron wind. Defects especially vacancies play a role, but grain boundaries and especially grain size play an even more significant role. In a perfect crystal, only volume and surface diffusion play a role, while polycrystallization promoting

grain boundary diffusion plays a much greater role. The electron wind preferentially separates atoms along grain boundaries where they are transported in the direction of the electron flow. For small grain sizes and a high interfacial fraction oriented in the direction of electron flow, electromigration is prominent. Reducing the grain boundary area or producing grain boundaries oriented perpendicular to the direction of the electron wind can significantly reduce electromigration effects. These features are illustrated in Fig. 5. Figure 5d, e illustrates the *bamboo structure* effect which is often a deleterious effect in other conducting wire systems (Fig. 5e) where grain boundary sliding can reduce the conducting cross section leading to *current crowding* and failure as in older incandescent light bulb filaments which failed after limited operation. Figure 5a–d illustrates on effective microstructure design strategy in metallization/interconnects where reduced interconnect sizes (areas) will become more conducive to orienting grain boundaries perpendicular to the electron wind direction and reducing electromigration. This occurs because there is a net reduction in grain boundary area and the corresponding internal energy of the metallization volume (or volume free energy).

Integrated Circuit Failure and Reliability

As integrated circuits become increasingly more complex with more components and decreasing feature sizes (Fig. 1), these components must become increasingly more reliable, thereby increasing or at least preserving acceptable reliabilities of the whole unit or system. As illustrated with only a few obvious examples of degradation and failure in Figs. 2, 3, 4, and 5, failure mechanisms in specific systems, chips, or other products become even more prominent at decreasing feature sizes which results in higher failure rates and shorter device lifetimes, along with unanticipated and often early device wear-out. The so-called physics-of-failure concepts and accurate mathematical approaches addressing or utilizing failure mechanism models can aid in reliability prediction. The failure rate contribution of each can be normalized by accounting for the effect of the weighted percentage of each failure rate. Simple approaches usually ignore the complex interactions between various failure mechanisms (Beck 1998; Wagner 1989; Ohring 1998).

Normalized failure rates can be defined for each ith failure mechanism as λ_i, and for each functional group type, the failure rate can be defined as a weighted summation of each failure rate type (λ_i) multiplied by a normalization constant for that specific failure mechanism, $K_{i,f}$ a constant defined by the weight percentage of a functional group, f, as it affects the ith failure mechanism:

$$\lambda'_f = \sum \left(K_{i,f} \times \lambda_i\right). \tag{1}$$

Correspondingly, the total failure rate of a component, λ_T, can be defined as being equal to the summation of the total number of each functional group multiplied by the failure rate of each functional group type:

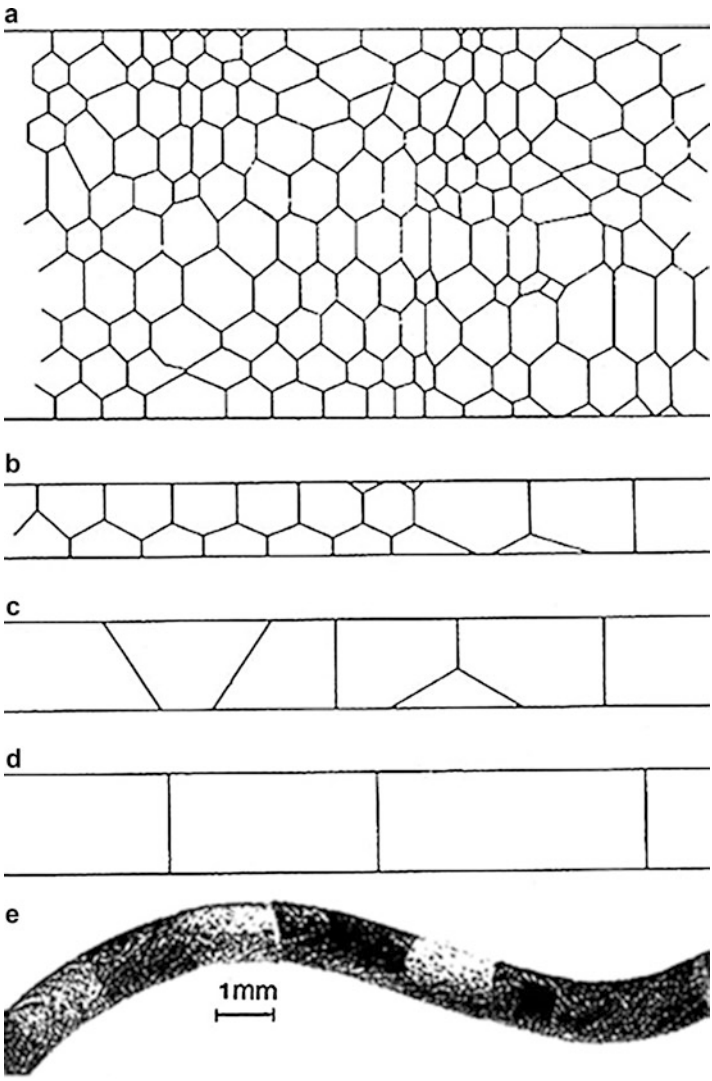


Fig. 5 Evolution of grain structure in metallic interconnect lines on ICs. (**a**) Many grains thick. (**b**) Thin interconnect line with few random grains creating a grain boundary diffusional path oriented in current (electron) direction. (**c**) Reduction of interior interconnect line grain boundary diffusion paths. (**d**) Creation of single-crystal (bamboo) segments with grain boundaries perpendicular to line surfaces. (**e**) Bamboo structure in Ti-Mo alloy wire

$$\lambda_T = \Sigma N_f \lambda_f = N \Sigma \left(\frac{N_f}{N}\right) \lambda_f = N \Sigma A_f \lambda_f, \quad (2)$$

where N_f is the total number of each functional group, N is the total number of all types of functional groups, and A_f is the ratio of the number of units of the nth

functional group type to the total number of functional groups existing in any specific component: $A_f = (N_f/N)$.

A rough estimate of the failure rate may be obtained from

$$\lambda = \left(\frac{\text{Total number of failures}}{\text{Cumulative working months}}\right). \tag{3}$$

Correspondingly, the mean time to failure (MTTF) is given by

$$\text{MTTF} = \int_o^\infty t f(t) dt = \int_o^\infty \lambda t e^{-\lambda t} = \frac{1}{\lambda}, \tag{4}$$

where f(t) is a probability density function and t is the time to failure. To describe wear-out,

$$f(t) = \left(\frac{1}{\sigma\, t\sqrt{\pi}}\right) \exp\left\{-\frac{1}{2}\left(\frac{\text{Int} - \text{Int}_{50}}{\sigma}\right)^2\right\} \tag{5}$$

where t_{50} is the median time to failure

$$\sigma \cong \ln\left(\frac{t_{50}}{t_{16}}\right), \tag{6}$$

with t_{16} being the time when 15.87 % of the device has failed.

For n products that fail after successively long times, t_1, t_2, t_3 . . .,

$$\text{MTTF} = \frac{(t_1 + t_2 + t_3 + \ldots . t_n)}{n}. \tag{7}$$

To illustrate how this analysis can be utilized, consider a total of 100 microcontrollers being replaced during a cumulative total of 500,000 working months. Equation 3 then becomes

$$\lambda = \frac{10^2}{5 \times 10^5} = \frac{0.2 \times 10^{-3} \text{failures}}{\text{month}}. \tag{8}$$

Correspondingly, the mean time to failure (MTTF) (Eq. 4) becomes

$$\text{MTTF} = \frac{1}{\lambda} = 5 \times 10^3 \text{months}. \tag{9}$$

It can be noted that the median time to failure, t_{50}, in Eq. 6 can also be defined as

$$t_{50} = \frac{\{(t_2 - t_1) + (t_3 - t_2) + \cdots (t_n - t_{n-1})\}}{n} \tag{10}$$

Equation 10 has also been known as Black's equation (Black 1999):

$$t_{50} = A\exp\frac{(E_A/kT)}{J^n}, \tag{11}$$

where A is a constant, E_A is the activation energy (~0.4–0.6 eV for Al; 0.8–0.9 eV for Cu), k is the Boltzmann constant, T is the absolute temperature, J is the current density, and n varies between $n = 1$ and $n = 3$. Equation 11 illustrates that failure times decrease strongly with increasing temperature and current density. Smaller and faster circuits cause higher current densities, lower voltage tolerances, and higher electric fields which will make future devices more vulnerable to early failure.

Accurately predicting failure rate can only become more difficult as feature size decreases and the transistors on a die continue to follow Moore's law, where the number of transistors doubles roughly every 2 years. Segregation of elements will at some point become impossible and cross sections of interconnects will become limited in current density. Although Al and W are prominent interconnect materials, Cu and even Ag are advantageous for higher conductivity and suppression of electromigration (Basaran and Lin 2007).

References

Basaran C, Lin MN (2007) Damage mechanics of electromigration in microelectronics copper interconnects. Int J Mater Struct Integrity 1:16–39

Beck F (1998) Integrated circuit failure analysis. Wiley, New York

Black JR (1999) Electromigration: a brief survey and some recent results. IEEE Trans Electron Dev 16:338–347

Ohring M (1998) Reliability and failure of electronic materials and devices. Academic, San Diego

Wagner LC (ed) (1989) Failure analysis of integrated circuits: tools and techniques. Springer Science and Business Media, LLC, New York

Part XV

Innovations in Magnetic, Electronic, Optoelectronic and Photonic Materials

Innovations in Magnetic Materials

Contents

Abstract

Magnetic physical principles and properties are reviewed to establish a basis for their application. Classifications of magnetic materials are presented as these apply to the role of magnetic dipole moments or magnetic polarization as it applies to magnetic memory and recording innovations and applications. The role of magnetic induction in power isolation and conversion is described, and functional and multifunctional magnetic materials innovations are discussed in terms of properties which include magnetoresistance, giant magnetoresistance, magnetostriction and magnetic shape-memory effects, magnetocaloric effect as it applies to magnetic refrigeration, and temperature effects on magnetic materials properties and performance.

Introduction: Review of Magnetic Physical Principles and Properties

It is notable that magnetism and magnetic materials have been of scientific interest for nearly a millennium, with many practical applications only realized within the past two centuries. Recent, fundamental investigations have been mainly

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_57

directed at exploring various types of magnetic materials and understanding novel magnetic effects and performance features created by electric currents.

The practical application of accumulated knowledge and understanding of magnetic materials has relied heavily on an understanding of the structure–properties–processing and performance relationships. The role of processing and its effect on magnetic materials microstructure evolution has had a significant effect on so-called extrinsic magnetic properties such as coercivity, remnant magnetization, and permeability.

In a historical perspective, the major applications of magnetic materials, apart from permanent magnets, has been in the magnetic recording area (Mee and Daniel 1996; Daniel et al. 1998), one of the most widespread magnetic materials applications area today, with an estimated annual data storage capacity of roughly 1 GB per year per person. Magnetic storage or recording applications involve particulate magnetic materials, magnetic thin film materials, along with multilayer forms. The discovery of *spintronics* or spin transport electronics (also known as magnetoelectronics), exploiting both the intrinsic spin of the electron as well as its associated magnetic moment and electronic charge in solid-state devices, has added a new dimension to the field of magnetic recording (Grünberg et al. 1986; Shinjo 2009). Stored magnetic information from a recording disk is read with the help of *giant magnetoresistance* (GMR) discovered in a magnetic multilayer system (Fe/Cr/Fe) around 1986. This historical breakthrough by Albert Fert (see Baibach et al. 1988) and Peter Grünberg (Grünberg et al. 1986) was awarded the 2007 Nobel Prize in Physics. In more conventional electronics, the charge of the electron is controlled and manipulated with the help of an electric field, while in spintronics there is an additional degree of freedom, the spin of the electron, which can be controlled by a magnetic field.

In addition to magnetic recording, another important area of magnetic materials application involves sensors, transducers, and actuators where the underlying property is *magnetostriction*, the most important magnetoelastic phenomenon exhibited by magnetic materials. Magnetostriction involves dimensional changes of a magnetic material by the application of a time-varying magnetic field, converting the electrical/magnetic energy to mechanical energy. Recent developments in the magnetoelastic regime have included a new class of materials known as ferromagnetic shape-memory alloys (Engdahl 2000). In addition, magnetostrictive transducers having compositions such as (Tb, Dy) Fe_2, and related rare-earth transition metal intermetallics can generate strains 100 times greater than traditional magnetostrictive materials and are referred to as having *giant magnetostrictive (GMR) properties*.

New applications and innovations of magnetic materials include magnetic cooling or magnetic refrigeration where a magnetic material such as $Gd_5Ge_2Si_2$ discovered by Pecharsky and Gschneidner (1997) possesses a *giant magnetocaloric effect (GMCE)*. In addition to exploiting specific heat changes in a material under the influence of a magnetic field, large changes in resistivity, called *colossal magnetoresistance (CMR)*, can result in the transition from a parent insulating magnetic material to a metallic ferromagnet. Several transition metal oxides undergo this kind of field-induced insulation-to-metal transition, resulting in CMR (Tokura 2000).

This brief overview of some of the emerging magnetic materials innovations has presented a range of magnetic properties and physical quantities which may require a brief review and clarification to allow a more in-depth discussion of more specific magnetic materials innovation concepts. These are presented in the following:

Magnetic Field Strength and Magnetomotive Force – A current (I) flowing in an infinitely long conductor generates a magnetic field strength, H, perpendicular to the direction of the current and oriented by the right-hand rule:

$$H = I/2\pi a, \tag{1}$$

where a is the radial distance from the infinite wire. For a current circulating in a single coil of wire of radius R, the magnetic field intensity at the center of the loop is given by

$$H = I/2R. \tag{2}$$

Correspondingly, an electric current circulating in a solenoid having N coils or turns of wire of radius, r per unit length, n, produces a magnetic field (or field strength) inside the solenoid given by

$$H = NI/2\pi r = nI, \tag{3}$$

where NI is termed the *magnetomotive force (mmf)*.

Magnetic Flux and Flux Density – A magnetic field in a vacuum is defined by

$$d\phi = B\,.\,ndA \tag{4}$$

where ϕ is the magnetic flux in area A (or dA), and B is the magnetic flux density:

$$B = \mu_o H, \tag{5}$$

where μ_o is the magnetic permeability of the vacuum (or free space) and equal to $4\pi \times 10^{-7}$ H.m^{-1} (Henry/meter). In a magnetic material, the magnetic flux density is given by

$$B = \mu H = \mu_o \mu_r H. \tag{6}$$

In Eq. 6, μ_r is referred to as the relative magnetic permeability of a material ($\mu_r = \mu/\mu_o$), while μ is the magnetic permeability, an intrinsic material property.

When a magnetic flux passes through N coils of a solenoid, it induces an *electromotive force* (emf) in the circuit. This is referred to as electromagnetic induction and is proportional to the rate of change of the magnetic flux:

$$\text{emf} = -N\partial\phi/\partial t = -NA\partial B/\partial t = -\mu_o NA\partial H/\partial t \tag{7}$$

It might be noted that while B, H, and μ have been indicated as scalar values, they are actually tensors as illustrated previously in Part II, chapter "► Electromagnetic Fundamentals."

Magnetic Dipole Moment, Magnetization, and Magnetic Susceptibility – It might be recalled from Maxwell's equations (▶ Eq. 36 in chapter "Electromagnetic Fundamentals" of Part II) that an asymmetry exists because of the absence of a corresponding magnetic charge (the monopole, which has never been observed). In spite of this discrepancy, a pair of magnetic poles (a dipole) can be defined as an infinitesimal current loop (or circular conductor) carrying a current and considered to represent the elementary unit of magnetism: a magnetic moment or dipole moment represented by a vector quantity perpendicular to the current loop plane:

$$\mathbf{M} = \mathrm{IA} \ = \mathbf{n}\mathrm{I}\ \mathrm{dA}, \tag{8}$$

where **n** is a unit vector normal to the area enclosed by the current loop and in the direction of **H**.

At the microscopic level, an atom or molecule subjected to a magnetic field, H, and corresponding magnetic induction, B, causes the magnetic spin of each electron to align along its direction:

$$\mathbf{m} = \eta\mathbf{B} = \mu_o\eta\mathbf{H}, \tag{9}$$

where η is the atomic or molecular magnetizability. Magnetization of the material is then expressed by

$$\mathbf{M} = \mathrm{N}\mathbf{m}, \tag{10}$$

where N is density or number of magnetic dipole moments per unit volume of material. When the material is subject to a magnetic field, **H**, the magnetization is given by

$$\mathbf{M} = \chi_m\mathbf{H}, \tag{11}$$

where χ_m is a dimensionless factor referred to as the *magnetic susceptibility* of the material. The overall magnetic induction of a material subjected to a magnetic field, **H**, becomes

$$\mathbf{B} = \mu_o\mathbf{H} + \mu_o\mathbf{M}. \tag{12}$$

Substituting Eq. 11, Eq. 12 becomes

$$\mathbf{B} = \mu_o\mathbf{H}(1 + \chi_m). \tag{13}$$

From Eq. 6:

$$\mathrm{M_r} = 1 + \chi_m; \mu = \mu_o(1 + \chi_m) \tag{14}$$

Magnetoresistance and Magnetostriction – An electric field will appear inside a conductor carrying a current when a magnetic field, H, is applied. This electric field (called the Hall field) will deflect the flow of electrons causing an increase in the

electrical resistance, R. *Magnetoresistance* is the change in the transverse electrical resistivity, ρ_t, of a conductor subjected to a static magnetic field and varies as a function of the square of the magnetic induction or magnetic flux density, B^2. It is maximum when the magnetic field is perpendicular to the current density vector. Relative changes in electrical resistivity and electrical resistance are equal only if the magnetostrictive properties are negligible.

Magnetostriction is a reversible dimensional change of a ferromagnetic material when magnetized. It is an elastic change due to the motion and rotation of Bloch walls or magnetic domain walls or interfaces. Spontaneous magnetostriction can occur at the Curie temperature due to magnetic moment reordering, while field-induced magnetostriction can occur by volume change with shape conservation, shape charge at constant volume, and on a change in magnetization by applying a tensile or compressive stress. Magnetostriction is a dimensionless coefficient of linear change denoted λ, a strain at a given magnetic field:

$$\lambda = (\Delta \ell/\ell)_H. \tag{15}$$

In fact, for spontaneous magnetostriction,

$$\lambda_o = \varepsilon/3 \tag{16}$$

where ϵ is the spontaneous strain. When a demagnetized ferromagnet is remagnetized until saturation, the fractional linear change at saturation along the direction of the field is

$$\lambda_o' = 2\varepsilon/3. \tag{17}$$

Figure 1 shows the magnetostriction of some polycrystalline, ferromagnetic metals as a function of a D.C. magnetic field. Magnetostriction in single crystalline materials is anisotropic, and strains will be different in different crystallographic directions.

Magnetocaloric Effect (MCE) – When a magnetic field is applied to a paramagnetic or ferromagnetic material, it aligns the spins of unpaired electrons and reduces the magnetic moment disorder. If this process occurs adiabatically (without exchange of heat with the surroundings), the material's entropy is unchanged, and the energy released by the magnetic moment alignment creates lattice disorder which manifests itself in phonon production which increases the material's temperature: heating by adiabatic magnetism. Correspondingly, magnetization of a material at a given temperature, and then thermally isolated, will lead to a decrease in temperature by adiabatic demagnetism. Removing the heat evolved by cooling when the magnetic field is switched off will allow the electron spins to revert to their initial, random orientation, increasing the entropy and cooling the material: cooling by adiabatic demagnetization or *magnetic refrigeration* (Fig. 2).

The cooling capacity, Q, of the magnetic material or the amount of heat that can be extracted from the cold end to the hot end of a refrigerator in an ideal thermodynamic cycle is given by integrating the magnetic entropy change:

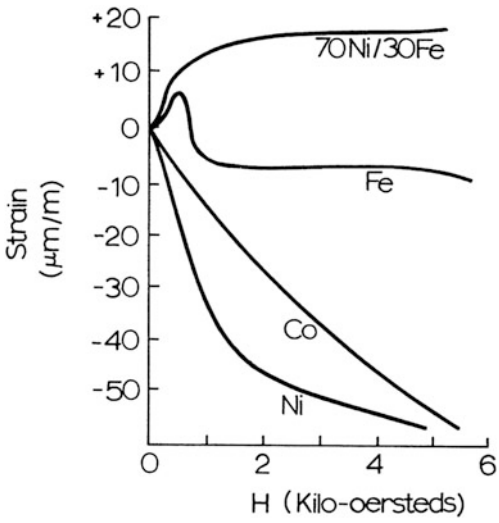


Fig. 1 Magnetostriction of ferromagnetic metals and alloys as a function of magnetizing field, H. Positive or negative strain corresponds to material expansion or contraction, respectively (Adapted from Murr (1978))

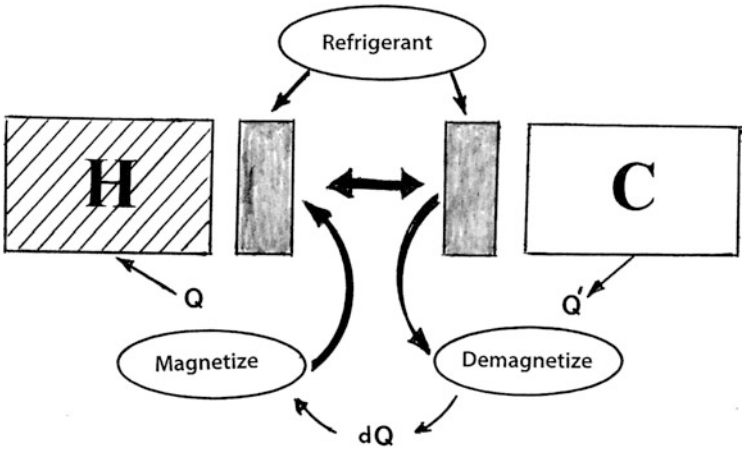


Fig. 2 Schematic of principal magnetic refrigerator components

$$dQ = -\Delta S_M(T)_{\Delta H} dT, \tag{18}$$

where $\Delta S_M(T)_{\Delta H}$ is the magnetic entropy change or field-induced isothermal magnetic entropy change at constant temperature per unit mass of material. Correspondingly, the adiabatic temperature drop or absolute temperature change of the material will be given by

$$(\partial T/\partial H)_S = -T/C_H(\partial M/\partial T)_H, \tag{19}$$

where C_H is the heat capacity and defined by

$$C_H = T(\partial S/\partial T)_H \tag{20}$$

in a constant magnetic field. As noted previously, a giant magnetocaloric effect (GMCE) has been observed in a Gd_5 (Ge_2 Si_2) alloy (Pecharsky and Gschneidner 1997) which exploits change of specific heat (or heat capacity) of a material.

Spintronics – Spin-polarized electron injection from a ferromagnetic metal to a normal metal (Johnson and Silsbee 1985) and giant magnetoresistance (Baibach et al. 1988; Grünberg et al. 1986) led to spin-dependent electron transport phenomena in solid-state devices: magnetoelectronics or spin transport electronics (Wolf et al. 2006; Shinjo 2009). Spin of an electron has an associated magnetic moment:

$$m_s = \sqrt{3}\, q\hbar/2m_o, \tag{21}$$

where q is the charge, ħ is Planck's constant (h) divided by 2π ($h/2\pi$), and m_o is the electron rest mass. In a spintronic device, generation or manipulation of a spin-polarized electron population results in an excess of electrons with their spins oriented in the same direction (either up or down).

A spin-polarized current can be generated in a metal by passing the current through a ferromagnetic material or a giant magnetoresistance (GMR) device. Such devices consist of at least two layers of ferromagnetic materials separated by a spacer layer. When the two magnetization vectors of the ferromagnetic layers are aligned, the electrical resistance will be lower and a higher current will flow at constant voltage in contrast to anti-aligned ferromagnetic layers. This arrangement constitutes a magnetic field sensor. Spin rotations or alignment in thin film sandwich arrays can also be utilized in more efficient data storage innovations to be discussed later.

Magneto-Optic Effect – A propagating electromagnetic wave such as light can have its plane of polarization rotated in a layer of magneto-optic material, often referred to as a gyromagnetic material which has been altered by a quasistatic magnetic field, either externally or internally because of intrinsic ferromagnetism. This causes a change in the permittivity tensor, ε, of the material (Part VI). This of course causes different refractive indices of the magneto-optic layer for left – and right – circularly polarized wave parts of the polarized light (the Faraday effect).

Magnetic Impedance – can be considered to be analogous to electrical impedance; the ratio of sinusoidal magnetic voltage to the sinusoidal magnetic current:

$$Z_M = V_m/I_M, \tag{22}$$

and

$$I = I_M \exp(-j\omega t). \tag{23}$$

When a soft ferromagnetic conductor is subjected to a small alternating current, a large change in the alternating current (AC) complex impedance of the conductor can be achieved on applying a magnetic field. This is known as the *giant magnetoimpedance (GMI) effect* (Phan and Peng 2008). Materials exhibiting this effect are used in magnetic sensor applications. In contrast to magnetoresistance

(MR), reflecting a change in resistance of a magnetic material subjected to a magnetic field, the magnetoimpedance (MI) effect consists of a change in the total impedance (including resistance and inductance) of a magnetic conductor under a direct current (DC) applied magnetic field, when a high-frequency alternating current follows. MR can contribute significantly to the MI effect at low frequencies but is negligible at high frequencies. GMI occurs in multilayer or sandwich thin films such as CoFeSi B/Cu/CoFeSiB, FeSiB/Cu (Ag)/FeSiB, and related arrays of ferromagnetic/highly conductive (nonmagnetic)/ferromagnetic (F/C/F) thin film sandwiches (Zhou et al 2001).

Classification of Magnetic Materials

Magnetic materials, those materials that respond to a magnetic field or are themselves magnetized owe their magnetic properties to the orbital motion and spin of electrons in an atom which create a magnetic dipole moment as expressed in Eq. 8. This concept of magnetic polarization is illustrated in Fig. 3a for an extracted volume element in a solid magnetic material having a relative permeability, μ_r. Consequently, the magnetic field intensity in the cavity volume is given by μ_r H, and the current, I, circulating in the cavity is given by

$$I = (\mu_r - 1)\mathbf{H} \cdot \mathbf{dL}. \tag{24}$$

Substituting Eq. 24 for I in Eq. 8 results in

$$\mathbf{m} = (\mu_r - 1)\,\mathbf{H}\,dL\,dA. \tag{25}$$

In some cases, magnetic polarization is described by magnetization or magnetization per unit volume of magnetic material:

$$\mathbf{M} = \mathbf{m}/dL\,dA = N\mathbf{m}, \tag{26}$$

where N is the number of dipole moments per unit volume (Eq. 10) or substituting from Eq. 25,

$$\mathbf{M} = (\mu_r - 1)\ \mathbf{H} = \chi_m \mathbf{H}. \tag{27}$$

Magnetic materials are distinguished to a large extent on the basis of their magnetization behavior and correspondingly can be characterized by their susceptibility. There are five classes of magnetic materials:

1. *Diamagnetic materials*
2. *Paramagnetic materials*
3. *Ferromagnetic materials*
4. *Ferrimagnetic materials*
5. *Antiferromagnetic materials*

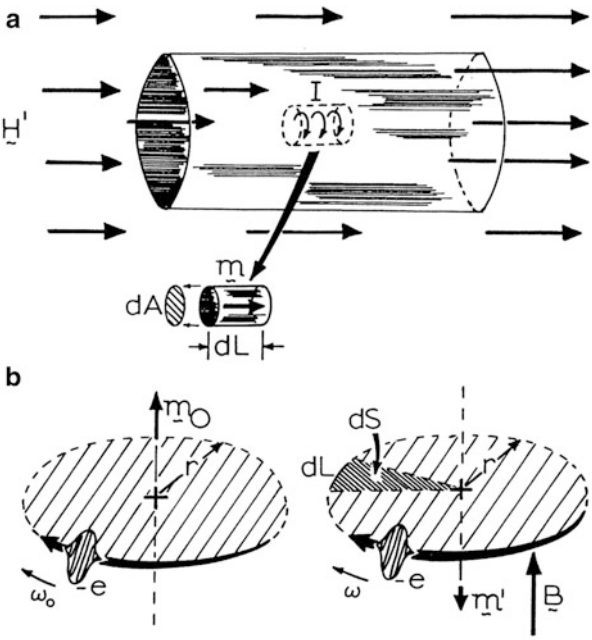


Fig. 3 Magnetic dipole (moment) convention (**a**) and atomic orbital magnetic dipole moment and induced orbital diamagnetic dipole moment (**b**). Dipole moment is designated **m**. Note field (H) current (I) direction by right-hand rule. In (**b**), the application of a magnetic field, **B** produces a reversed dipole moment, $\mathbf{m}'$ in contrast to the intrinsic, orbital magnetic dipole moment, $\mathbf{m}_o$ at *left* (Adapted from Murr (1978))

It might be recalled from chapter "▶ Structures and Properties of Oxide Superconductors," and especially Fig. 1, that superconductivity or the superconducting state represents a state of perfect diamagnetism. Figure 3b illustrates this concept for an electron orbiting a nucleus in a single atom where on the left the orbital magnetic dipole moment associated with this orbital electron motion is given by Eq. 8 as

$$\mathbf{m}_o = \mathbf{n} I_e A_o = -\bar{e}\omega_o r^2 \mathbf{n}/2, \tag{28}$$

where I_e is the current generated by the electron motion:

$$I_e = -\bar{e}\,\omega_o/2\pi, \tag{29}$$

where A_o is the orbital area ($A_o = \pi r^2$ in Fig. 3b), $\bar{e}$ is the electron charge, and ω_o is the angular frequency of the electron. When a magnetic field, **B**, is applied as in the right of Fig. 3b, an electric field, E, is created in the circular electron orbit:

$$E = -\frac{r}{2}\left(\frac{dB}{dt}\right). \tag{30}$$

This will cause a force on the orbital electron: $F = -eE$ which will produce a change in angular momentum by

$$m_o r d\omega = F dt, \tag{31}$$

where m_o is the electron rest mass. Substituting Eq. 30 into Eq. 31 and integrating results in

$$\omega = \omega_o + \overline{e}B/2\, m_o. \tag{32}$$

Substituting Eq. 32 into Eq. 28 results in

$$\mathbf{m}' = -\mathbf{n}\left(\frac{\overline{e}r}{2}\right)^2\left(\frac{B}{m_o}\right), \tag{33}$$

where the negative sign indicates that, as shown in Fig. 3b, this dipole moment is opposite in direction to the applied magnetic flux density, **B**, and the orbital magnetic dipole moment, $\mathbf{m}_o$ (not to be confused with the electron rest mass). Since Fig. 3b represents a single dipole moment per atom, N in Eq. 26 could represent the number of atoms per unit volume of solid. In a general case of an atom with atomic number Z (or Z electrons) having a spherically symmetric charge with mean square radius $= 2r^2/3$, the corresponding magnetic (or diamagnetic) susceptibility can be written as

$$\chi_m{}^{(dia)} = -2\mu NZ\overline{e}^2r^2/18\, m_o, \tag{34}$$

where it can be noted that the sign of the susceptibility is negative. Normal diamagnetic materials (solid materials) are Be, B, Ga, In, Si, Ge, Pb, Cu, Ag, and Au among other elements, such as C in the form of diamonds and ionic compounds such as MgO and Al_2O_3. Organic materials are generally diamagnetic. However, perfect diamagnetic materials include all type I superconductors (chapter "▶ Structures and Properties of Oxide Superconductors").

To a large extent, diamagnetism is characterized by the relative magnitudes of the intrinsic, orbital magnetic dipole moment, $\mathbf{m}_o$, and the induced magnetic dipole moment, $\mathbf{m}'$, when a magnetic field is applied. For cases where the magnitude of $\mathbf{m}_o$ exceeds that of the induced magnetic dipole moment, $\mathbf{m}'$, the solid material will possess a net permanent magnetic moment which will always be in excess of the induced moment, $\mathbf{m}'$. When exposed to a magnetic field, this material will exhibit a small net magnetization in the direction of the applied magnetic field, **B**, and will be characterized as a *paramagnetic material*. Consequently, in contrast to diamagnetic materials, paramagnetic materials exhibit permanent dipole moments. However, in the absence of a magnetic field, the dipole moments are randomly oriented so that the net magnetization is zero. But since this randomness is a characteristic of each atom and its position in the lattice, temperature, as it affects lattice vibrations, will influence this randomness, even in a magnetic field. In other words, thermal disorder arising from lattice vibration will tend to resist magnetic dipole moment orientation in a magnetic field, and at very high temperature for any crystalline solid, the magnetization will be destroyed ($M = 0$). There is also a temperature where a material's permanent magnetism changes to induced magnetism. This is referred to as the *Curie temperature*, T_c, or Curie point. In other words, below T_c, magnetic dipole moments are aligned and random above. Consequently, a material

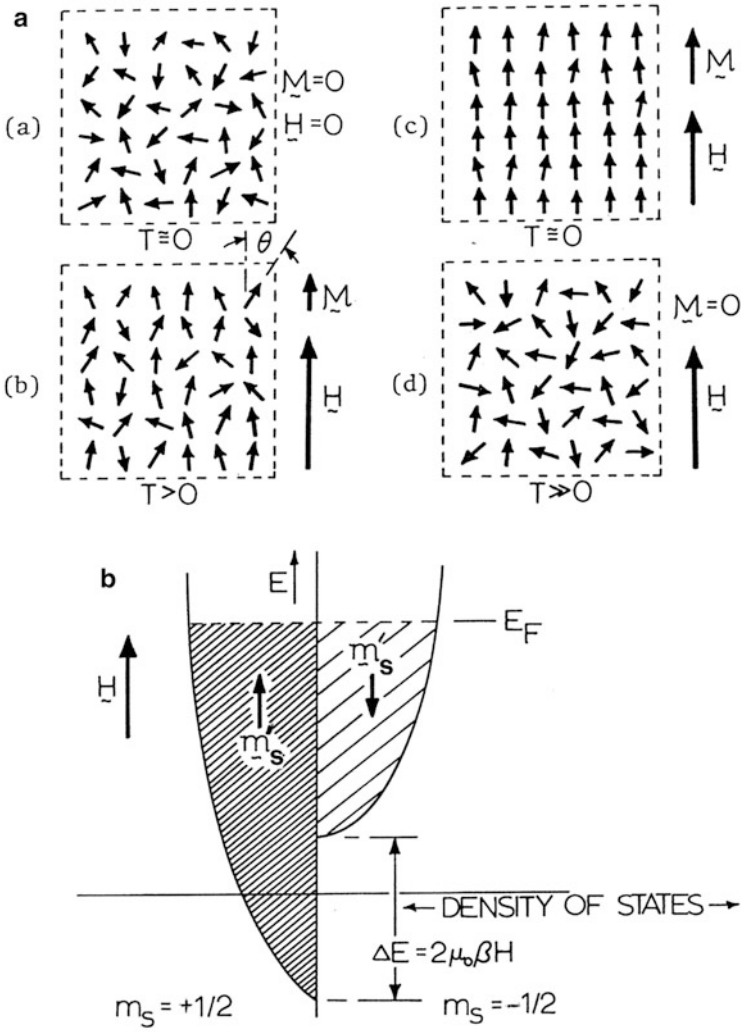


Fig. 4 Magnetism in paramagnetic materials. (**a**–**d**) show magnetic dipole moment orientation and corresponding magnetization (**M**) with an applied external magnetic field strength, **H**, as a function of temperature, T. (**e**) Free-electron paramagnetism at $T = 0K$. $\mathbf{m}_e$ is the magnetic dipole moment of spinning electrons (free electrons). m_s is the spin quantum number ($\pm$½) as shown

is paramagnetic only above its Curie temperature and magnetic only when a magnetic field is applied. These features are generally illustrated in terms of the magnetization as represented in Fig. 4a. The net magnetization, M, in the direction of the applied magnetic field intensity, H, is given by

$$M = Nm^2\mu_o H/3K\,(T - T_c), \tag{35}$$

Table 1 Magnetic susceptibilities for diamagnetic and paramagnetic materials (in cgs units at room temperature)[a]

Diamagnetic materials		Paramagnetic materials	
Material	$\chi_m^{(dia)}$ (10^{-6})	Material	$\chi_m^{(para)}$ (10^{-6})
CuS	−2	TiO_2	5.9
Si	−4	Al	17
Cu	−5	Cr	180
C (diamond)	−6	V	200
Ag	−20	Mn	530
ZnS	−25	Pd	570
Al_2O_3	−37	V_2O_5	2,000
SiO_2	−43	FeO	3,600
In	−64	Er	4×10^4
Sb	−99	Er_2O_3	7.3×10^4
CuO	−240		

[a]See Cardarelli (2005): χ(cgs emu) = 4πχ(SI units)

and from Eq. 26, the paramagnetic susceptibility becomes

$$X_m(\text{para}) = Nm^2\mu_o/3K\,(T - T_c) \equiv e/(T - T_c), \tag{36}$$

where C is the paramagnetic Curie constant = $(Nm^2\mu_o/3K)$ and K is Boltzmann's constant. In contrast to diamagnetic susceptibility, paramagnetic susceptibility is positive. Table 1 compares examples of diamagnetic and paramagnetic materials.

While consideration of magnetization in paramagnetic materials as illustrated in Fig. 4a has only included orbital electron magnetic moments as shown in Fig. 3b, free or unbound electrons can also interact with a magnetic field, thereby contributing to the magnetization and the paramagnetic susceptibility. These electrons can be represented by the conduction band or energy level configurations shown in ▶ Fig. 5 of chapter "Electromagnetic Color and Color in Materials" or a schematic view representing two halves of the conduction band to account for electron spin values, m_s (spin quantum number having a magnitude ± ½), as shown in Fig. 4b. There are slightly more electrons in the conduction band half representing spins $m_s = 1/2$, with a magnetic moment of the spinning electron represented by $\mathbf{m}_s'$, which is parallel to the applied field, **H**, represented by ΔE in Fig. 4b. The density of electrons with magnetic dipole moments in the field direction can be approximated as the product of ΔE and the density of states at the Fermi level (E_F). Consequently, the net magnetization is given by

$$M = 2N(E_F)\,\beta^2\mu_o H, \tag{37}$$

where

$$\mathbf{m}_s' \cdot \mathbf{n} = \pm\beta, \tag{37a}$$

and β is a magnetic quantum called the Bohr magneton ($\beta = 9.273 \times 10^{-24}$ joule/Wb/m^2; see Cardarelli (2005) for conversion of units). **n** in Eq. 37 is a unit vector. The magnetic moment, $\mathbf{m}_s'$, also has an interaction energy with the magnetic field defined by

$$E_S = \mu_o \mathbf{m}_s' \cdot \mathbf{H} \tag{38}$$

or

$$E_S = \pm \mu_o \beta H; \tag{39}$$

the electron spin interaction energy or exchange energy, ΔE, in Fig. 4b.

From Eq. 26, the electron paramagnetic susceptibility component becomes

$$\underset{\text{electron}}{\chi_m^{(\text{para})}} = 2\beta^2 N(E_F)\mu_o, \tag{40}$$

where $N(E_F)$ is the electron density of states at the Fermi level. Depending upon the density of states or the prominence of free electrons in a solid material, the electron paramagnetic susceptibility may be negligible. In addition, and as illustrated in Eq. 40, it is not temperature dependent.

The concept of an exchange interaction implicit in Fig. 4b embodies the internal interaction between the localized magnetic moments, $\mathbf{m}_s'$. This occurs due to the *Pauli Exclusion Principle:* two electrons having opposite (antisymmetrical) spins can occupy the same orbital or angular momentum state (or quantum substate, S^2) and be closer to each other, thereby producing a strong Coulomb repulsion. Correspondingly, electrons with the same spin (symmetrical) will occupy different orbitals or quantum substates, producing a smaller Coulomb repulsion because they are further apart. This is illustrated in Fig. 5. The exchange energy or the energy due to the repulsion between the two electrons is thereby minimized, and the Coulomb repulsion force favors the parallel alignment of all electron spins as the exchange energy is minimized. This can be observed more specifically for iron in the context of the electronic structure for iron, in particular the 3d energy-bond or quantum substate structure. Figure 6a shows the absence of an exchange interaction where there is no net spin magnetic moment. Figure 6b on the other hand shows a magnetized state which possesses the lowest total energy and correspondingly the most stable state in Fe, which is representative of a ferromagnetic material. In ferromagnetic materials like Fe, Co, and Ni, magnetic dipoles are aligned parallel to one another as shown in Fig. 6c even without an external magnetic field, producing a net or spontaneous magnetization, **M**, or $\mathbf{M}_s$, where

$$\mathbf{M}_s = N\mathbf{m}_s, \tag{41}$$

and N is the number of atoms per unit volume, and the internal magnetic field associated with this magnetization as shown in Fig. 6b is

$$\mathbf{H}_E = \lambda \mathbf{M}_s, \tag{42}$$

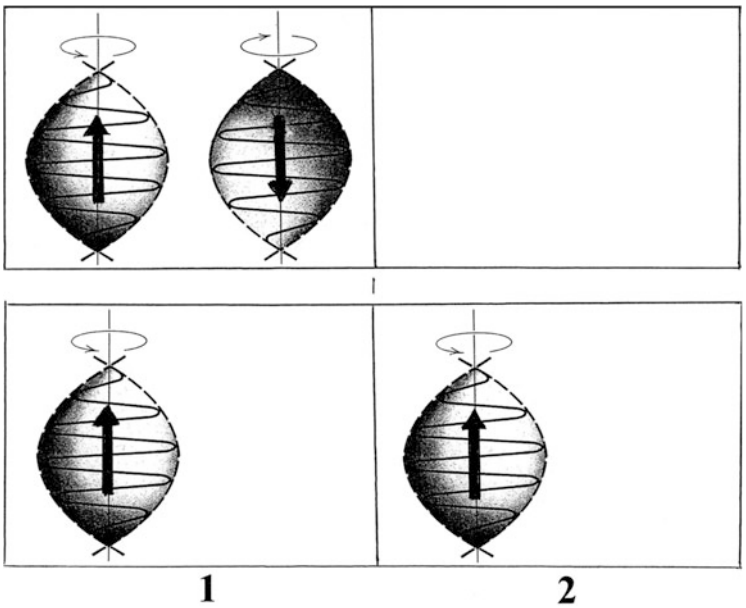


Fig. 5 Schematic showing two electrons occupying orbital (*1*) producing a strong Coulomb repulsion (*top*) compared to two electrons occupying different orbitals (*1*) and (*2*) (*bottom*) having a smaller Coulomb repulsion, minimizing the exchange energy. Note spin moments (*arrows*)

where λ is referred to as the Weiss field constant after P. Weiss, who first postulated the existence of $\mathbf{H}_E$ in 1907. With the application of an external magnetic field (intensity), **H**, each magnetic dipole will experience a total field

$$\mathbf{H}_T = \mathbf{H} + \lambda \mathbf{M}_s, \tag{43}$$

and substituting the scalar equivalent of Eq. 43 into Eq. 35 with $\mathbf{H} = \mathbf{H}_T = \mathbf{H} + \lambda \mathbf{M}_s$, the ferromagnetic susceptibility becomes

$$\chi_m(\text{ferro}) = C/\left[T - \left(N\mu_o m^2 \lambda/3K\right)\right]. \tag{44}$$

However, the Curie temperature can be set equivalent to the second term in the denominator of Eq. 44:

$$T_c = N\mu_o m^2 \lambda/3K, \tag{45}$$

where upon Eq. 44 becomes

$$\chi_m(\text{ferro}) = C/(T - T_c); \tag{46}$$

referred to as the Curie–Weiss law for ferromagnetic materials. In this case, implicit in Eq. 46, ferromagnetic materials lose their spontaneous magnetization and become paramagnetic above the Curie temperature, T_c ($T \geqq T_c$).

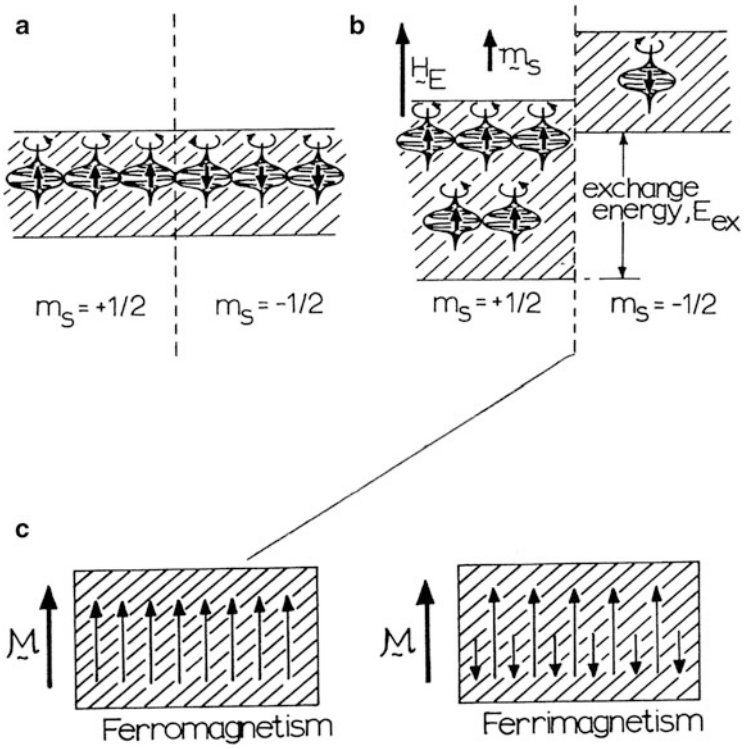


Fig. 6 3d energy-band schematic for ferromagnetic iron in unmagnetized state (**a**) and magnetized state (**b**) giving rise to a spontaneous spin moment, $\mathbf{m}_s$, for each atom experiencing an external magnetic field strength, $\mathbf{H}_E$. (**c**) Compares spin (dipole moment) ordering with magnetization in ferromagnetic and ferrimagnetic materials (Adapted from Murr (1978))

As illustrated in Fig. 6d, *ferrimagnetic materials* have atoms with unequal spins oriented in an antiparallel fashion. In this case, the spontaneous magnetization can be regarded as the two opposing and unequal magnetizations of the atoms on two sublattices. Like ferromagnetic materials, ferrimagnetic materials become paramagnetic above the Curie temperature, corresponding phenomenologically to Eq. 46.

Antiferromagnetic materials have an antiparallel arrangement of equal electron spins which results in a very low magnetic susceptibility similar to paramagnetic materials expressed in Eq. 36. This phenomenon is observed primarily in ionic compounds with the NaCl crystal structure. The spin arrangement of antiferromagnetic materials becomes unstable above a critical temperature called the Neel temperature, T_N. Figure 7 shows the prominent features of antiferromagnetism in MnO which has the NaCl structure shown in Fig. 7a. Figure 7b shows the susceptibility versus temperature response where the shaded portion indicates the region in which the spins (Fig. 7a) are antiparallel as shown in Fig. 7d in contrast to ferromagnetism shown in Fig. 7c. The susceptibility of antiferromagnetic materials in the paramagnetic region to the right of T_N in Fig. 7b ($T > T_N$) is given by

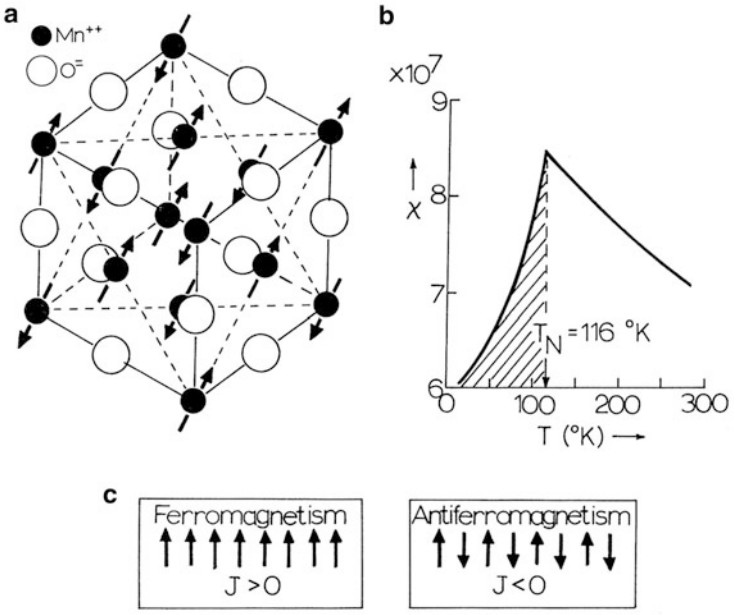


Fig. 7 Antiferromagnetism in M_nO. (**a**) Antiparallel spin orientations (*arrows*) of M_n^{2+} ions in M_nO NaCl-type unit cell. The {111} planes shown *dotted* exhibit antiparallel spins. (**b**) M_nO magnetic susceptibility versus temperature. (**c**) Comparison of spin (dipole moment) ordering in ferromagnetism in contrast to an antiferromagnetic material like M_nO. J is the exchange integral (a coefficient) (After Murr 1978)

$$\chi_m(\text{ferri} - \text{para}) = 2\,C/(T + T_N), \tag{47}$$

where C is the Curie constant.

Magnetic Domains and Domain Walls in Ferromagnetic Materials

Even in single crystal ferromagnetic materials, magnetic dipole alignment is not usually continuous. Regions in which the dipole moments are aligned are referred to as domains or magnetic domains. While it might appear that continuous alignment of magnetic dipoles in a single domain would represent energy minimization in the context of the exchange interaction represented in Fig. 6b, there are other contributions to the total magnetic energy. These include *magnetostatic energy*, *magnetostrictive energy*, and *magnetocrystalline energy*.

Figure 8 illustrates magnetostatic energy and its reduction by forming magnetic domains separated by domain walls or boundaries. The demagnetizing field has a magnetostatic energy which depends on sample shape and allows work to be done by the magnetized sample. By dividing the material into domains as shown in Fig. 8, the external demagnetizing field is reduced or eliminated,

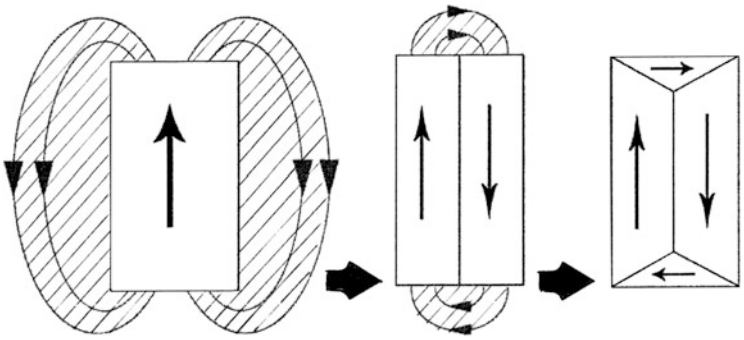


Fig. 8 Magnetostatic energy. *Shading* around the magnetized domains shows demagnetizing field which has a magnetostatic energy which is reduced as the sample is divided into magnetized domains. The magnetostatic energy is zero at the right since there is no demagnetizing field

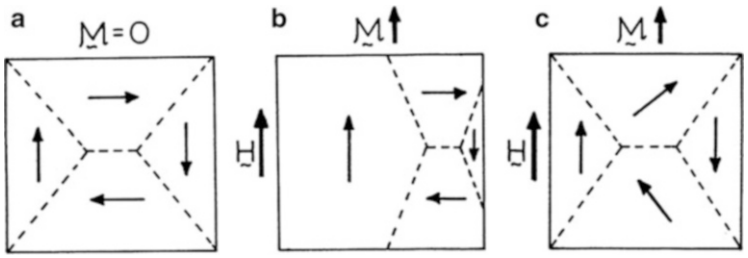


Fig. 9 Ferromagnetic domain concept. (**a**) Zero magnetization. (**b**) Domain growth in a magnetic field, **H**, produces net magnetization, **M**. (**c**) Domain (magnetization direction) rotation producing net magnetization, **M**. *Dotted lines* designate domain boundaries (After Murr (1978))

but correspondingly, the exchange energy increases. Nonetheless, the total energy has decreased since the magnetostatic energy dominates and is reduced or eliminated.

As illustrated in Fig. 1, magnetization of ferromagnetic materials changes their dimensions: magnetostriction. As shown in Fig. 1, these shape changes can be positive or negative (expansion or contraction). The corresponding elastic strain energy is proportional to the volume of the domains of closure. Consequently, magnostrictive energy is minimized by decreasing the domain size or increasing the number of domains. This adds domain walls which contribute to the magnetocrystalline energy and the exchange energy (Fig. 6b), which favors wide walls where adjacent magnetic dipole moments are as close to parallel as possible. Figure 9 illustrates the ideal ferromagnetic domain concept as it relates to magnetization, **M**, while Fig. 10 illustrates simple domain wall concepts. Figure 10a shows Bloch walls where the spin rotations occur so that a 180° twist in direction occurs as the domain wall is crossed. The angular twist between spins of adjacent atoms is denoted ϕ, and the domain wall width is W_D. The atom spacing or lattice spacing is denoted a. Figure 10b shows Neel walls where the 180° change in

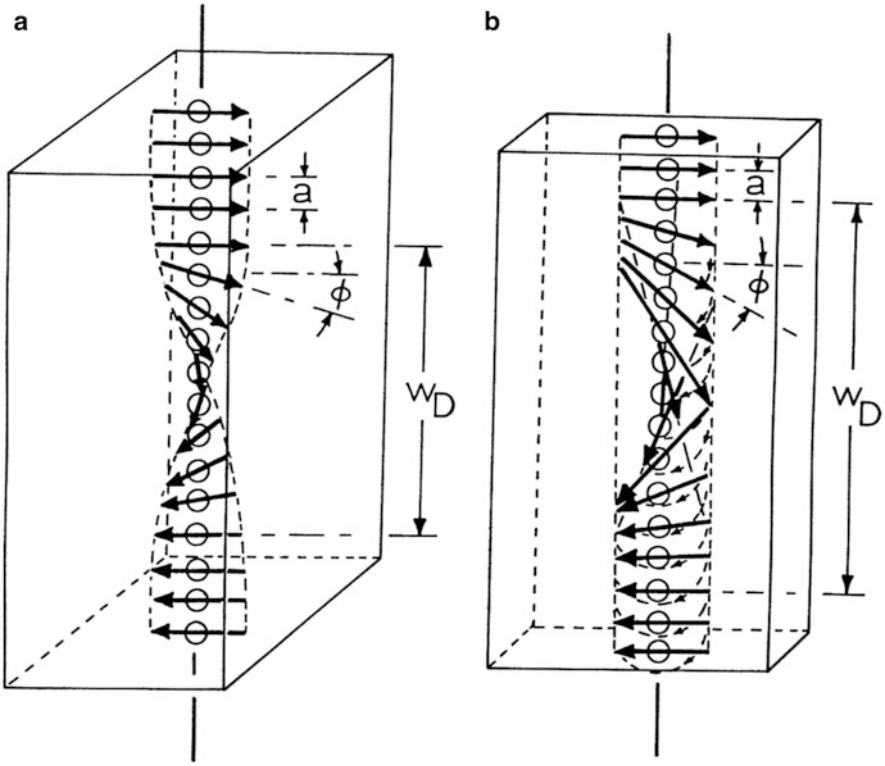


Fig. 10 Domain wall concepts. (**a**) Bloch wall section in a ferromagnetic material observed along a single line of atoms in a plane of the lattice, spaced *a*. (**b**) Neel wall in a thin ferromagnetic section of material observed along a line of atoms spaced *a*. The domain walls are denoted W_D (Adapted from Murr (1978))

magnetization direction occurs by rotation in a plane. Bloch walls usually have a lower energy than Neel walls and are predominant except in thin film ferromagnetic materials where the magnetic dipole moments rotate around an axis perpendicular to the thin film surface plane.

The domain wall energy per unit area of Bloch wall can be approximated using a simple cubic material in the following form (Fig. 10a):

$$E_D \cong JW_D\phi/2a^3, \tag{48}$$

where J is the exchange integral which also characterizes the exchange energy (Fig. 6b):

$$E_{ex} = Jm_s{}^2\phi^2, \tag{49}$$

where ϕ represents the same angle between spin moments as shown in Fig. 10a.

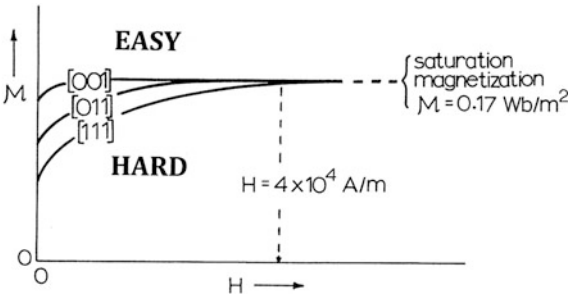


Fig. 11 Magnetic anisotropy (crystal anisotropy) for iron in an external magnetic field. EASY and HARD magnetization directions are noted as <001> and <111> directions, respectively (Adapted from Murr (1978))

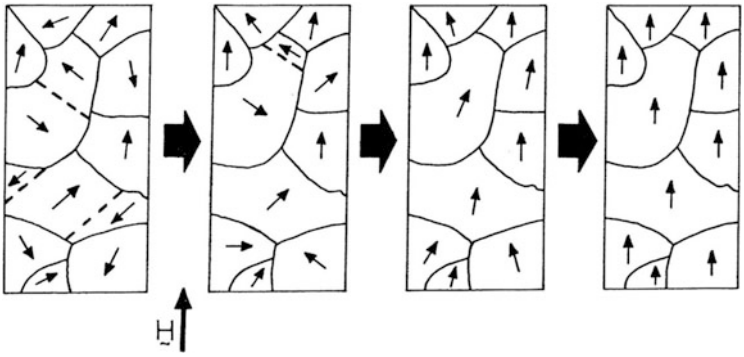


Fig. 12 Polycrystalline ferromagnetic material magnetization in a magnetic field, **H**. Grain boundaries are represented by *solid lines*, magnetic domain boundaries by dashed lines. *Arrows* indicate magnetization directions of individual crystal grains or magnetic domains (After Murr (1978))

Like many other phenomena in crystalline materials, magnetization, or the orientation of magnetic dipole moments, is also influenced by crystal directions. In cubic materials such as Fe (bcc) or Ni (fcc), there are "easy" and "hard" crystal directions which correspond to a lower or higher energy or energy difference associated with magnetization. This corresponds to the magnetocrystalline energy. The domain walls and in particular the domain wall energy also contributes to the magnetocrystalline energy. Consequently large domains and few domain walls minimize the magnetocrystalline energy.

Figure 11 illustrates the magnetic anisotropy associated with "easy" and "hard" magnetization directions in crystalline Fe. For fcc Ni, the magnetic anisotropy is similar, but the "easy" direction in Ni is <111>, while the "hard" direction is <001>. As shown in Fig. 11, <011> is an intermediate magnetization direction. This is also the intermediate magnetization direction in fcc Ni as well. For Co (hcp), the "easy" magnetization direction is <0001> parallel to the hcp c-axis.

Magnetocrystalline energy is higher for polycrystalline materials because grain boundaries contribute to additional anisotropy energy. Crystal defects can also add small increments as well since they can interact with the domain walls. As illustrated in Fig. 12, magnetization in polycrystalline ferromagnetic materials involves

spin moment or magnetic dipole moment rotations within each grain, and the movement and annihilation of domain walls in larger grains until saturation magnetization is reached at the far right. It can be observed in this schematic that for sufficiently small (nano-) grain sizes, single grains constitute single magnetic domains. In this context, it can be observed that for some applications involving speciation of magnetic dipole moment orientation or magnetization, nanocrystalline materials or nanostructures pose unique magnetic properties. Correspondingly, amorphous magnetic (ferromagnetic) materials are also observed to possess unique magnetization features.

Magnetic Hysteresis

Magnetic hysteresis refers to the irreversibility of the magnetization and demagnetization process. Hysteresis is embodied in a B–H magnetization curve and hysteresis loop or diagram as illustrated in Fig. 13a. The magnetic induction starts at Eq. 1 in Fig. 13a from zero at zero magnetic field ($B = 0$ and $H = 0$) gradually increasing the magnetic field. During this stage, magnetic domains, having the same spontaneous magnetization oriented with the applied field, will grow by motion of the domain walls. These walls will encounter defects in the crystalline grains and become pinned. Increasing the field frees the pinned walls which grow into a single domain having a net maximum induction value called the saturation magnetic induction, B_S, corresponding to a maximum value of the magnetic field strength, H_S (at Eq. 2 in Fig. 13a). Decreasing the magnetic field follows a different curve with higher values of magnetic field strength, H, because of the irreversibility of the domain wall motion. At zero field ($H = 0$), the magnetic domains reappear slowly, and a remnant magnetic induction or residual magnetic field remains in the material. Applying an opposite magnetic field can remove the retained magnetism or demagnetize the material. This opposite field is called *coercivity* or coercive magnetic field strength, H_c. As the opposite magnetic field is reduced to $H = 0$, the remnant magnetic induction, B_R, characterizes the residual magnetic induction. The coercive field is the field needed to reduce the magnetization to zero from saturation. Remagnetization by reversing the magnetic field from B_S at Eq. 5 in Fig. 13a completes the B–H curve by going through Eqs. 6 and 7 in Fig. 13a to saturation at Eq. 2 ($+B_S$) in Fig. 13a. The area bounded by the hysteresis curve or loop represents the stored magnetic energy per unit volume of material or the magnetic energy loss (hysteresis loss) per unit volume. These curves also illustrate the characteristics of magnetization/demagnetization. Considering that magnetization points up in the upper portion and down in the lower portion of the hysteresis loop, it illustrates the magnetic switching response of a particular ferromagnetic material, including the energy to switch. Figure 13b compares hysteresis loops for domain wall motion on the left and rotation on the right, indicating less energy required (or lost) on the right. Similarly, Fig. 13c shows a material on the left which, like Fig. 13b (left), requires a large field to magnetic and demagnetize. Such materials are referred to as *hard ferromagnetic materials* and exhibit a large

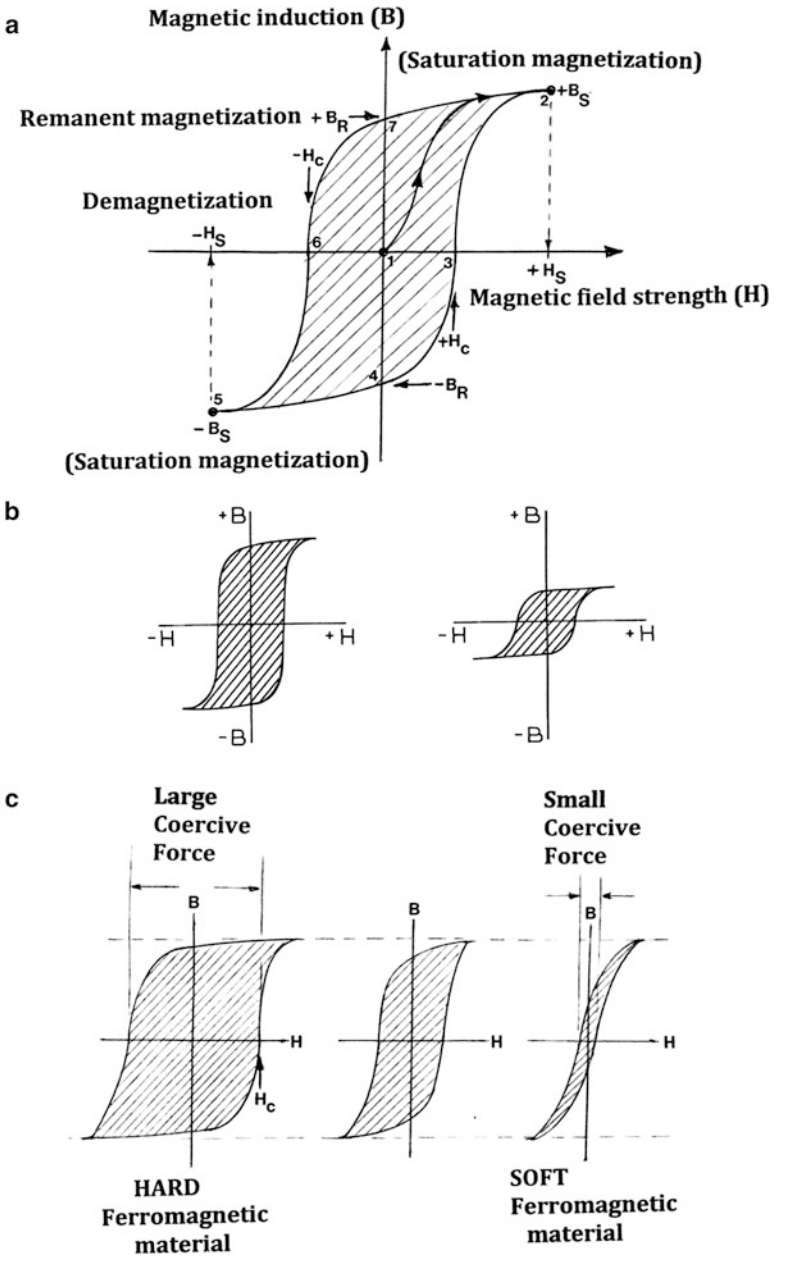


Fig. 13 B–H hysteresis loop (diagram) showing magnetization – demagnetization path enclosing magnetic energy (*BH*): shaded (**a**). (**b**) Compares hysteresis loops for polycrystalline magnetic materials exhibiting preferential domain wall motion (*left*) and rotation (*right*). (**c**) Sequence of hysteresis loops comparing HARD (*right*) and SOFT (*left*) ferromagnetic materials. Center diagram is intermediate. H_c is the coercive field

coercive force in contrast to *soft ferromagnetic materials* at the right in Fig. 13c which exhibit a small coercive force. Hard magnets are referred to as permanent magnets and have a large remnant magnetization. Soft magnets, however, are easily magnetized and demagnetized (or switched). The hysteresis diagrams in Fig. 13b and c illustrate a range of ferromagnetic performance features which have been adapted to a wide range of ferromagnetic materials applications and innovations over many decades. It might also be noted that magnetic permeability of ferromagnetic materials depends on a particular region of the B–H diagram. Consequently, it is not a constant physical parameter and therefore not a useful parameter in characterizing ferromagnetic materials. A more useful physical quantity is the so-called differential permeability:

$$\mu = \partial B / \partial H. \tag{50}$$

Soft magnetic materials having low coercivity or a small coercive force as shown in Fig. 13c exhibit high permeabilities. Owing to these high permeabilities, soft ferromagnetic materials can be magnetized with low to moderate magnetic fields and retain only a residual magnetic induction when the magnetic field is switched off. Table 2 compares some soft magnetic metals and alloys with hard magnetic alloys.

Electromagnetic Devices: Power Isolation and Conversion Applications

Historically, power generation and conversion, including motors covering a wide range of sizes and power requirements, and an equally large range of transformers, including even micro-transformers on chips, have accounted for the largest segment of magnetic applications, including hard, permanent magnetic materials as well as soft magnetic materials applications which include a wide array of solenoids and related electromagnetic devices. Among important hard and soft magnetic materials have been ferrites (spinel structures) discussed previously in chapter "▶ Electrovalent Crystal Structures and Chemistry," Part IV. These have a basis in spinel crystal structures and are mixed oxide derivatives with a basic Fe_2O_3 formula unit or derivatives of magnetite, Fe_3O_4 ($Fe^{2+}Fe_2^{3+}O_4^{-2}$: AB_2O_4). The spinel structure accounts for semiconducting and insulating behavior, representing resistivities ranging from $\rho \cong 10^{-2}$ to 10^{12} Ω-cm, especially useful at very high-frequency operation where low loss is important. The hysteretic behavior of ferrites also makes them particularly attractive in high-speed switching applications.

Voltage or current conversion (up or down) is performed by transformer configurations where an input or primary winding (P) or induction loop induces a time-varying magnetic flux which links an output or secondary winding (S), inducing in it a voltage or current determined by Faraday's law (▶ Fig. 13 and Eq. 27 in chapter "Electromagnetic Fundamentals"). The electromotive force (emf) of self-induction in a coil carrying current can be expressed as

Table 2 Properties of select hard and soft magnetic metals and alloys

Hard metal or alloy	Composition (wt.%)	Curie temp. Tc (°C)	Remnant magnetization B_R (T)[a]	Coercivity H_c (kA/m)[a]	Soft metal or alloy	Composition (wt.%)	Curie temp. T_c (°C)	Saturation magnetization B_S (T)[a]	Coercivity H_c (A/m)[a]
Alnico 5 (cast)	51 Fe – 8.5 Al-14.5 Ni – 24 Co – 3 Cu	860	1.27	51	Cobalt	Co	1,130	1.78	–
Alnico 8 (cast)	34 Fe – 7 Al – 15 Ni – 35 Co – 4 Cu – 5 Ti	860	8.25	131	Ferrosilicon	96 Fe-4 Si	735	1.97	24
Alnico 12 (cast)	33 Fe – 6 Al – 18 Ni – 35 Co – 8 Ti	860	0.60	64	Armco Iron	99 Fe	770	2.16	32–80
Chromindur II	Fe-28 Cr-10.5 Co	630	0.98	32	Metglass	81 Fe-20B	–	1.60	3.2–6.4
Cr-Steel	Fe – 3.5 Cr	745	0.95	5.3	Permalloy 45	55 Fe-45Ni	480	1.58	24
Co-Sm	Sm Co_5	725	0.90	640	Permendur	50 Fe-50Co	980	2.46	160
Ferrite 1 (sintered)	BaO – 6 Fe_2 O_3	450	0.22	145	Perminvar	43Ni-34Fe-23Co	–	1.50	2.4
Ferrite 4 (sintered)	SrO – 6 Fe_2O_3	450	0.40	175,000	Rhometal	64 Fe-36Ni	275	1.00	39.8
Co-steel	Fe-36 Co-3.75 W – 5.75 Cr – 0.8 C	890	0.98	19	Sendust	85 Fe-10 Si-5Al	480	1.00	3.98
Nd-27	Nd Fe B	280	1.08	740	Steel	Fe-0.8 C	–	2.15	144

[a]See Cardarelli (2005)

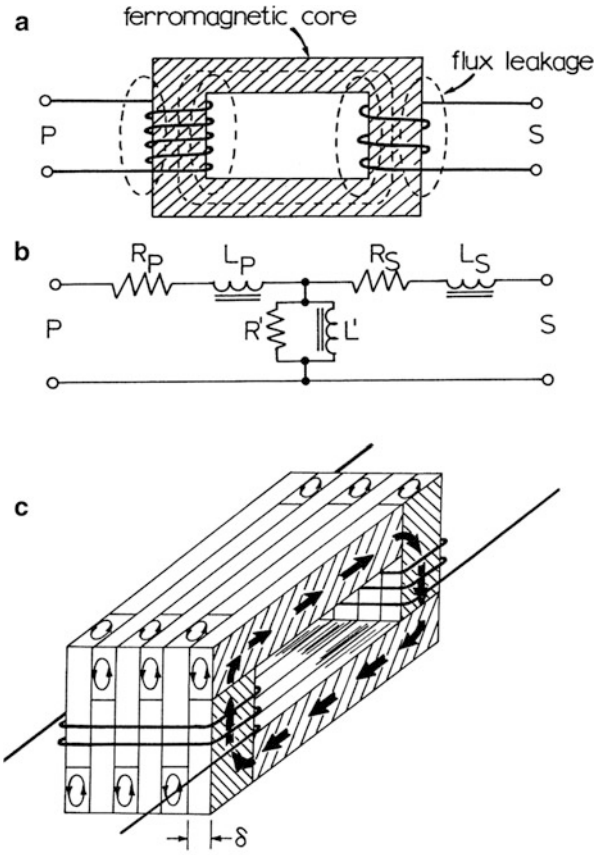


Fig. 14 Ideal transformer schematics. (**a**) and (**b**) show transformer components and equivalent circuit diagram, respectively. R_p, L_p, R_s, L_s denote equivalent winding (primary, P, and secondary, S, resistance and inductance, respectively). R′ and L′ are mutual resistance and inductance arising from magnetic coupling between the primary and secondary windings. (**c**) Laminated (magnetic) core transformer. *Arrows* in laminations indicate eddy-current circuits. Heavy arrows represent the magnetic field (**B**) direction in the core (Adapted from Murr (1978))

$$e_L = -N\frac{d\phi}{dt} = -\frac{d(\phi)}{dt} = -L\frac{di}{dt}, \tag{51}$$

where L is the self-inductance, and i is the instantaneous current in the circuit; ϕ is the magnetic flux which links N turns of wire coils. Considering the simple, ideal transformer shown in Fig. 14a and b, the emf of self-induction in either the primary (P) or secondary (S) windings will be

$$e_P = -N_P\frac{d\phi}{dt}; e_S = -N_S\frac{d\phi}{dt}. \tag{52}$$

Since the line voltage in the primary and secondary windings will be equal in magnitude but opposite in direction to the induced emf, Eq. 52 can be written as

$$V_P = N_P\frac{d\phi}{dt}; V_S = N_S\frac{d\phi}{dt}, \tag{53}$$

and since dϕ/dt is constant in the primary (P) and secondary (S) windings

$$V_P/V_S = N_P/N_S. \tag{54}$$

Correspondingly, the relationship between primary (P) and secondary (S) current will be

$$I_P/I_S = -N_S/N_P, \tag{55}$$

where the negative sign on the right side of Eq. 55 indicates that the currents in the primary and secondary windings (N_P and N_S) are opposite in phase. It can be noted in Eqs. 54 and 55 that while the voltage is being stepped up or down, the current is being correspondingly stepped down or up. That is, voltage and current in a circuit transform or convert in an opposite sense (Fig. 14a and b). Ideal transformers or transformer constriction is illustrated in Fig. 14c. The key issue involves a laminated core of magnetic material which can be selected on the basis of requisite resistivities, coercivities, and thermal dissipation of electrical energy in the core, which is frequency dependent as implicit in equations of the form

$$P_c = (\pi \ell f B_m)^2/6\rho, \tag{56}$$

where P_c is referred to as eddy-current (Foucault current) power loss (power released, as Joule heating per unit mass of magnetic material) in ferromagnetic laminates or sheets of thickness, with current flow at a frequency, f (Hz) a magnetic induction of B_m, and an electrical resistance, ρ. It can be observed from Eq. 56 that eddy-current power loss is minimized at low frequency operation with very thin ferromagnetic material laminates separated by an insulating layer, having high resistivity.

There is also a hysteresis heat loss or energy loss per cycle (f) in the form

$$P_h = fV_cA_{hys}, \tag{57}$$

where V_c is the total volume of core material, and A_{hys} is the area enclosed by the material's hysteresis loop as shown in Fig. 13. Since the area depends on the coercivity, H_c, which is roughly proportional to the grain size of the magnetic material, the heat loss is reduced in nanomaterials or especially amorphous materials where the resistivity, ρ, is also proportional to the reciprocal grain size. Conversely, the total power loss:

$$P_{TOT} = P_e + P_h \tag{58}$$

can be minimized by using nanograin laminates or amorphous (or glassy) alloy laminates. This is especially tractable for smaller transformer configurations where amorphous ribbon can be used effectively, but not necessarily in large transformers which are mostly limited to small grain permanent magnetic alloys such as FeSi (McHenry et al. 1999).

While surface-mount electronic technologies can incorporate small transformers characterized by very small wire-wound configurations shown in Fig. 14, shrinking electronic device configurations and integrated circuit (IC) configurations (on chips (dies)) require miniaturized power converters or micro-inductor or

micro-transformer configurations. While these can include air or nonmagnetic core configurations (Meyer et al. 2010) as is possible with larger-scale air core transformers, the inductance is limited, and these configurations are limited to lower operating frequencies.

Novel configurations utilizing MEMS processing, 3D printing, and standard CMOS-IC processing have produced *on-chip micro-transformers* as illustrated in Fig. 15a and b (Yunas et al. 2009; Yunas and Majlis 2008). Single deposition of the magnetic core material provides a completely closed magnetic circuit and minimizes leakage flux. However, in this approach, it is difficult to introduce uniaxial anisotropy in the core which is a necessary condition for high-frequency operation. As in other devices and especially IC components, the areal density or footprint area of the micro-transformer is an important consideration, and areas <2 mm^2 are achievable with small conductor widths <30 μm in planar coils shown in Fig. 15.

While structures like Fig. 15a can be fabricated by ion or laser-beam micromachining of bulk (micro-bulk) components similar to MEMS (microelectromechanical systems) fabrication, these features can also be fabricated on-chip using 3D printing technologies discussed in chapter "► Bioprinting and Biofabrication of Organs," Part XI. Figure 15b shows a deposition-etching sequence which also produces stacked, interwinding transformer configurations. Similarly, Fig. 15c and d illustrate CMOS-IC-type layer fabrication of an inductor element (Fig. 15c) and a micro-transformer (Fig. 15d), where the primary or secondary (input–output) are arbitrary for step-up or step-down function. The configurations in Fig. 15c and d can also be fabricated by 3D printing processes as well and are intended primarily as conceptual illustrations.

In contrast to micro-transformers illustrated in Fig. 15a, Fig. 16 illustrates one of the largest power transformers. The contrast between Figs. 15a and 16 is an example and testament to the vast spectrum of applications of materials science and engineering and especially the realm of magnetic materials and magnetic materials innovations.

Magnetic Recording and Memory Systems and Devices

Magnetic recording, storage, or memory systems are essentially synonymous since in all cases data is stored in patterns of magnetization in a magnetizable material. These patterns are simply oriented dipole moments which in the most recognizable data formats are opposite in direction, plus or minus, which can correspond to zero (0) or one (1) in a binary (2) system. In these systems, data can be deposited and retrieved in magnetized domains which retain the data or magnetization patterns indefinitely or until erased and rerecorded. In digital audio or video systems, the fundamental binary (0 or 1) system can also allow storage and retrieval. However, analog signals can also be embedded in a magnetization distribution which can be read out to reproduce the original signal.

The earliest memory utilized small magnetic toroids or magnetic rings which embodied Ampere's law implicit in ► Fig. 3a in chapter "Electromagnetic

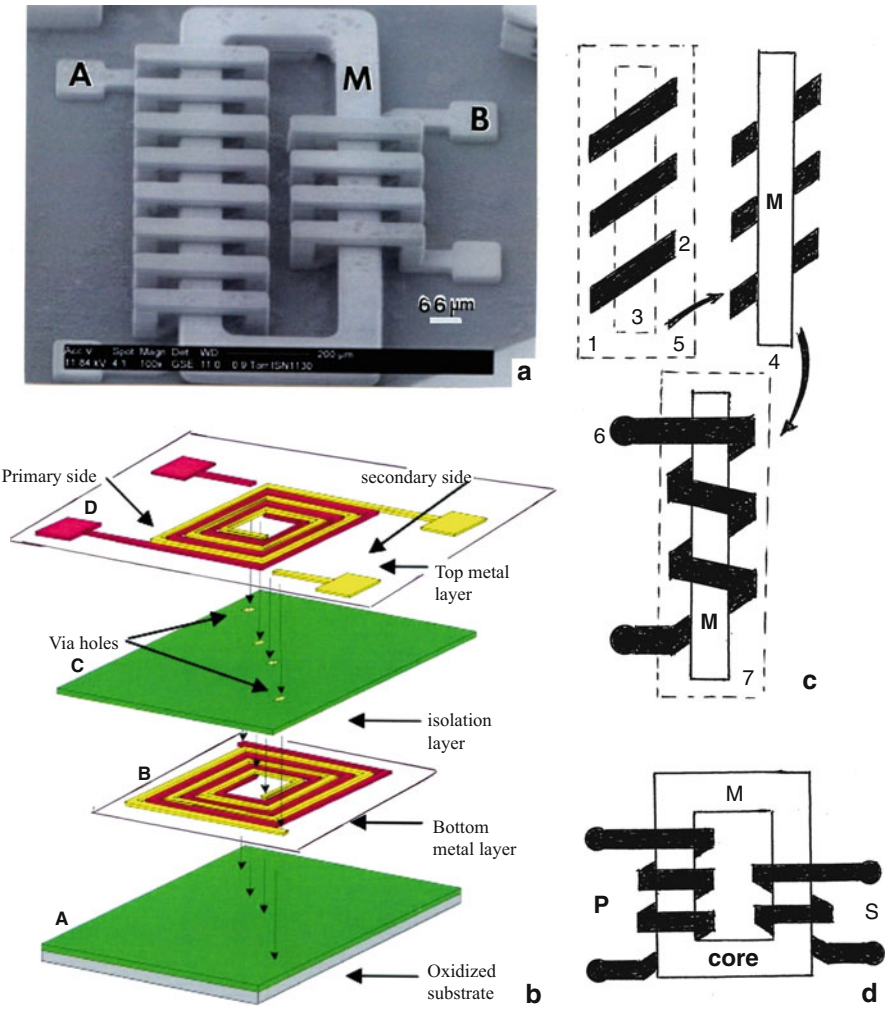


Fig. 15 Micro-inductor/transformer fabrication. (**a**) SEM image showing 3D micro-transformer fabricated on-chip by layer-by-layer (additive manufacturing or 3D-printing) (Courtesy of MEM Gen Corp., Burbank, CA). *A* and *B* denote input or output leads (primary or secondary). M is a ferromagnetic core material. (**b**) Stack-interwinding IC process fabrication of micro-transformer on-chip. Interlayer etching creates planar coil structure and via holes (Adapted from Yunas and Majlis (2008)). (**c**) CMOS-IC fabrication of layered inductor. SiO_2 insulator layer (*1*) followed by metal conductors (*2*). SiO_2 insulator layer (*3*), ferromagnetic (M) material layer (*4*). SiO_2 overlayer (*5*), coil completion (metal) layer (*6*), and SiO_2 insulation overlayer (*7*). (**d**) Same sequence as (**c**) to create on-chip micro-transformer

Fundamentals" or the right-hand rule as illustrated in Fig 17a. By creating a system of wires threaded through these magnetic rings or cores as illustrated in Fig. 17b, the cores could store one bit of data as shown in Fig. 17a as either a zero or one (0 or 1) simply using two drive lines (x, y) corresponding to A, a, B, b. . . . etc. in Fig. 17b.

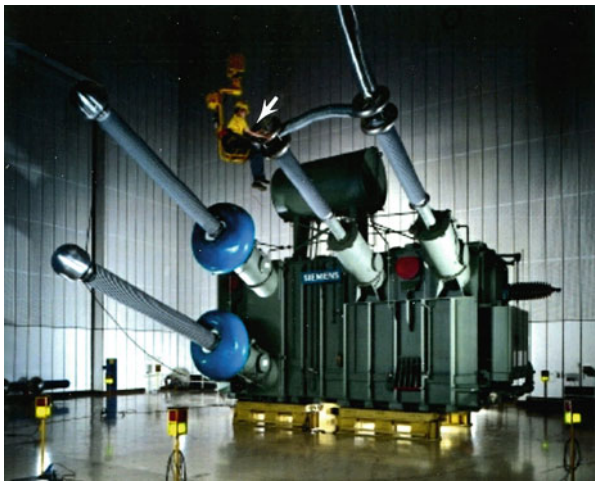


Fig. 16 Example of one of the largest electrical power transformer configurations. 800 kV DC (HVDC), 235 MVA transformer (Courtesy Siemens AG)

In principle, the two drive lines threading a magnetic core can produce four combinations of current direction: two combinations produce a reinforcing field in the core, while two other combinations cancel the fields in the core. This is illustrated in the sequence shown in Fig. 17c. The cores can also be arranged as shown in Fig. 17d where the two drive lines are rearranged so that one drive line passes through both cores in opposite directions creating binary combinations of 0 and 1. As illustrated in Fig. 17b, one sense loop goes through all cores, while other wires may also be used to reset the memory (or the state of magnetization). Interrogation of the cores requires separate switching matrices. For an $n \times n$ memory (Fig. 17b is a 3×3 memory), there must be two essentially identical switching matrices each of size $\sqrt{n} \times \sqrt{n}$. Consequently, an original matrix of n elements requires a switching matrix of elements.

Cores for early memory were sintered ferrites which were fabricated into $n \times n$ memory modules or 3D – $n \times n \times n$ memory modules as illustrated in Fig. 18a and b, respectively. Cores in early computer memory arrays in the early 1950s were around 2 mm in diameter and shrank to ~0.4 mm in diameter by the early 1970s. Correspondingly, access speeds of ~200 kHz increased to >1 MHz. Memory modules with ~500,000 cores were manufactured in the 1970s, although 3D memory arrays of several million bits were described by J.W. Forrester in the early 1950s (Forrester 1951). Costs of as much as one dollar per core (or bit) declined to 0.01 dollars per bit between the 1950s and 1970s. However, even at core sizes of 0.4 mm, the memory module was large and could be accommodated in the physical structures of large, mainframe computers. By the end of the 1970s, magnetic core RAM (random access memory) was superseded by semiconductor (transistor) RAM memory, which could be more easily fabricated especially in more dense (areal) chip arrays (Reilly 2003). Figure 19 illustrates a comparison of magnetic core memory arrays overlain by an encapsulated memory chip array: 8 bytes versus 8 GB (gigabytes) (1 byte = 8 bits).

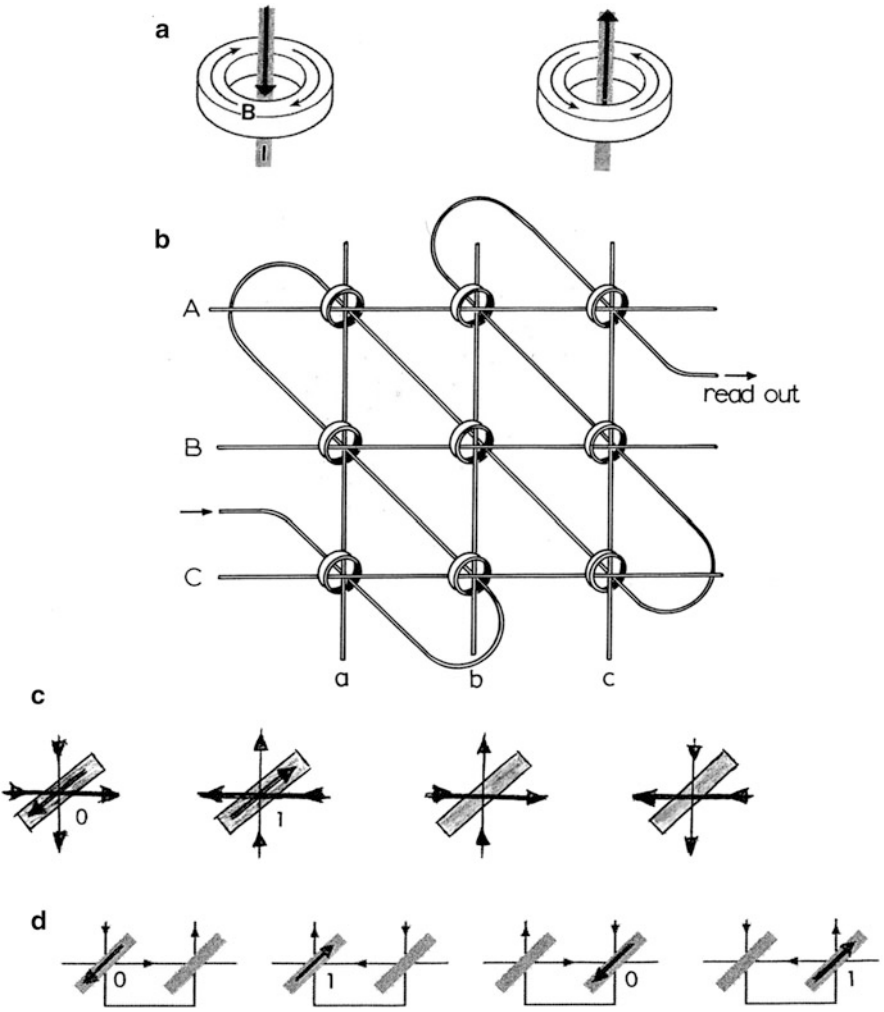


Fig. 17 Magnetic core memory principles. (**a**) Ampere's low (right-hand rule): (*left*) thumb (current) down, induced field (*B*) clockwise (fingers). (*right*) Current up, field counterclockwise. (**b**) 3 × 3 ferrite core memory array with two current drive lines. (**c**) Two states (0 or 1). Two states blank (no effect). (**d**) Drive line passes through both cores in opposite directions to produce all possible binary (0 or 1) combinations (**b** is adapted from Murr (1978))

Long before magnetic core memory and its use in computer memory arrays, a machine for recording speech magnetically (or by magnetization pattern development) on steel wire, called the telegraphone, was demonstrated by V. Poulsen, a Danish engineer, in 1900. During World War II, German engineers developed an audio tape recording system called the magnetophone which was the precursor to a wide range of tape recording and memory systems: data storage and retrieval. A magnetic tape recording/memory system in its simplest form is illustrated in

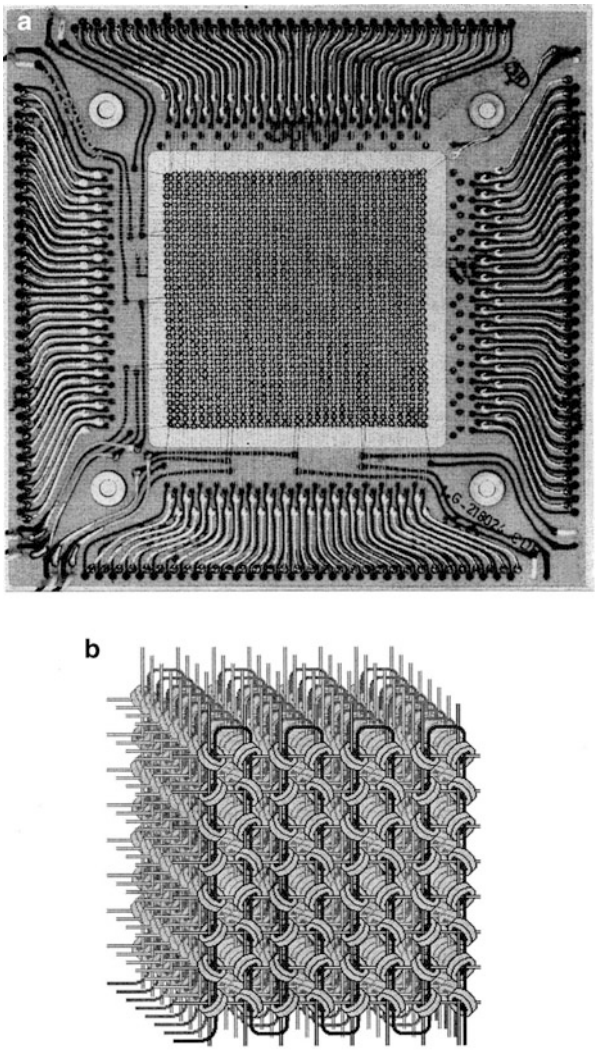


Fig. 18 Magnetic (ferrite) core memory. (**a**) 32 × 32 core array (~5 cm on a side). (**b**) 8 × 8 × 8 3D core array concept. Wikimedia Commons

Fig. 20a. Input signals in audio, video, or data (analog or digital; a.c. or d.c. as in Fig. 20b) in the form of a localized magnetic field with controllable intensity are created in the gap in the core winding. A magnetic tape, which can take many forms of magnetizable material, is then moved with constant velocity past the gap (or head as it is called) which creates magnetic dipole moment configurations characteristic of the input signal. The magnetization can be played back or retrieved by creation of an open-circuit emf in the winding as the tape moves past the head at the same velocity as the recording velocity. Convenient and cheap tape configurations have been, and continue to be, fabricated by placing ferrite (Fe_3O_4) and other magnetizable oxides (γ-Fe_2O_3, Cr_2O_3, etc.) in a suitable binder which is coated over a

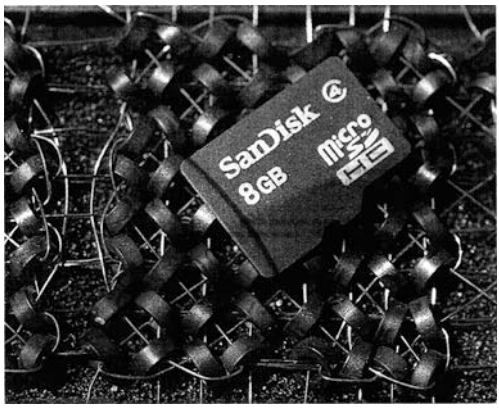


Fig. 19 1960s magnetic core 8×8 or 8 byte array overlain by contemporary (micro SDHC card) electronic memory of 8 GB (gigabyte). Ferrite cores are ~8 mm diameter. Wikimedia Commons (SDHC: Secure Digital High Capacity)

polymer base as illustrated in Fig. 20c. These polymer-based tapes have the convenience of being easily spooled and can be made in various tape widths. Magnetic tape strips continue to be used on a variety of credit card and related plastic commerce card formats, but microchips and other 2D bar code (or quick response code, QR) information records are becoming more convenient and reliable. In the case of powder crystal recording media (Fig. 20c), randomly oriented crystals can be appropriately magnetized as shown schematically in Fig. 20a. It might be noted that tape recording and memory formats are referred to as planar or *longitudinal magnetization formats* in contrast to simply orienting magnetic dipole moments up or down in a *perpendicular magnetization format* which will be discussed later.

A more popular device for storing and retrieving digital information has been the hard disk drive (HDD) which is a rotating disk coated with magnetic material as in Fig. 20c. In this configuration, data stored in a magnetized medium is read in a random access manner: magnetic random access memory (MRAM). In other words, individual blocks or zones which can correspond to magnetized bits representing 0 or 1 can be stored or retrieved in any order rather than sequential as in tape formats (Fig. 20a). The magnetic record–read head is arranged on a moving actuator arm. The disk, containing a 10–20 nm thick magnetic material (crystalline powder in a binder) can be ceramic, glass, or metal (usually an aluminum alloy). An outer layer of carbon or vapor-deposited graphite protects the magnetic material coating. Disks are rotated at speeds varying from ~4,200 rpm to 15,000 rpm in high-performance computing systems. To eliminate variances in magnetization of bit regions due to thermal instability, disks are coated with two layers of magnetic (or ferromagnetic) material, separated by a few atom layers of nonmagnetic ruthenium metal, and the layers are magnetized in opposite directions, reinforcing each other.

Of particular interest in magnetic recording and especially in HDD technology is the memory density or data storage capacity, which can be translated as the areal density of bit regions (0 or 1 : + or −), measured in bits/in^2. Like memory in solid-state devices (SSDs) (chips), the memory density is of increasing concern, and HDD memory capacity for current disk sizes (diameters) of 1.8–3.5 in. (4.6–8.9 cm) has exceeded 1 TB (terabytes), corresponding to ~700 G bit/in^2 at the end of 2013

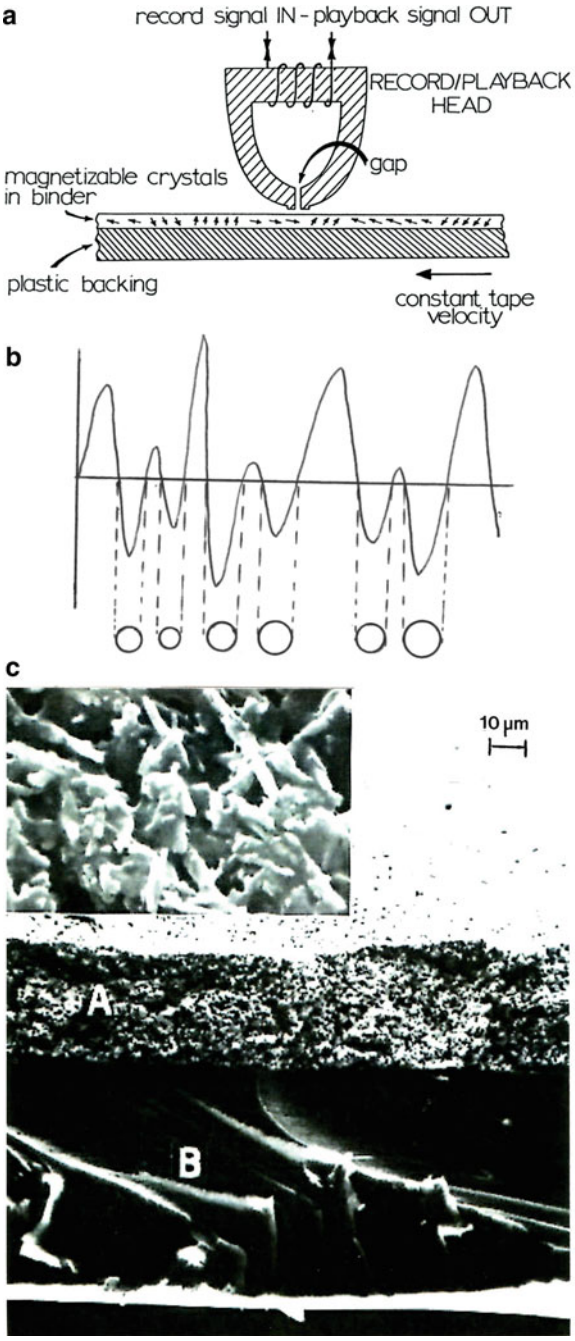


Fig. 20 Simple magnetic recording. (**a**) Record–playback in magnetic tape. (**b**) Analog-to-digital conversion using size and spacing encoding. (**c**) Magnetic tape cross section observed in the SEM. *A* is γ-Fe_2O_3 crystal particles in acetate binder. *B* is a plastic (polymer) backing. Insert shows magnetic crystal particles magnified 20 times the tape image marker shown. (**a** and **b**) are adapted from Murr (1978)

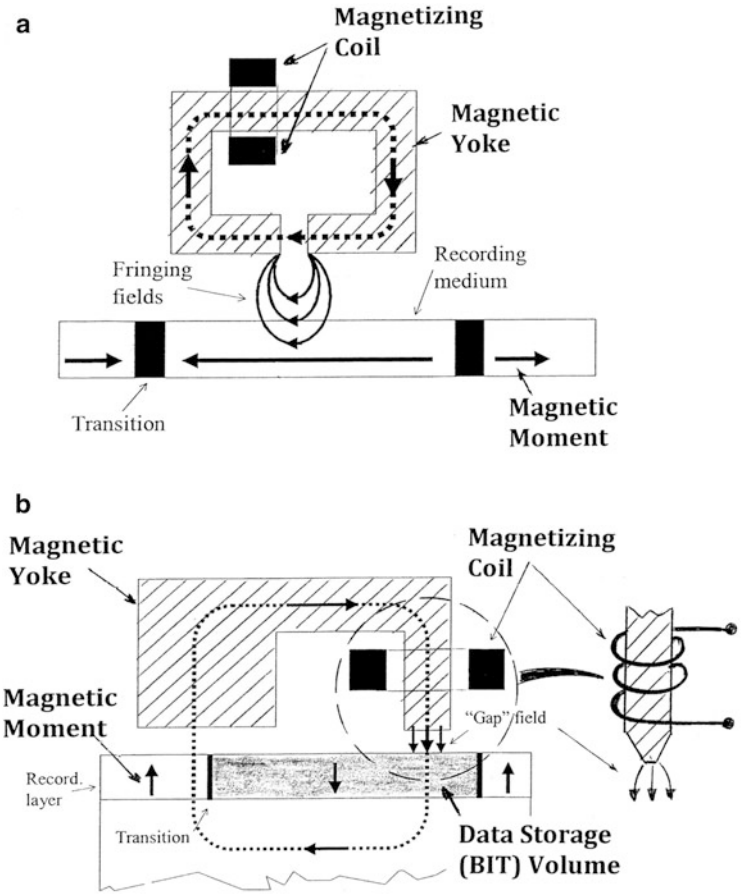


Fig. 21 Schematic diagrams comparing cross sections for longitudinal (horizontal) (**a**) and vertical or perpendicular (**b**) magnetic recording systems. Note concentrating record/retrieve head in (**b**) for small area bit recording (Adapted from Khizroev and Litvinov (2005))

and growing at a slowing annual rate of ~30 % because the memory capacity is nearing current technology saturation. This is due mostly to reconfiguring the magnetizing material bit regions so that the magnetization or dipole moments are oriented vertically or perpendicular to the disk surface: *perpendicular magnetization* or *perpendicular magnetic recording* (Khizroev and Litvinov 2005). Figure 21b illustrates this concept in contrast to longitudinal magnetic recording shown in Fig. 21a, which is conceptually the same as Fig. 20a.

Hard drives in small personal computers use giant magnetoresistance (GMR) sensors as a read head (Fig. 21b) which consists of a three-layer sandwich of a magnetic metal such as Fe or Co with a nonmagnetic metal filling such as Ag or Pt. Current perpendicular magnetic recording media consists of CoCrPt–SiO_2 nanogranular films in which nanosized, columnar CoCrPt grains with hcp structure

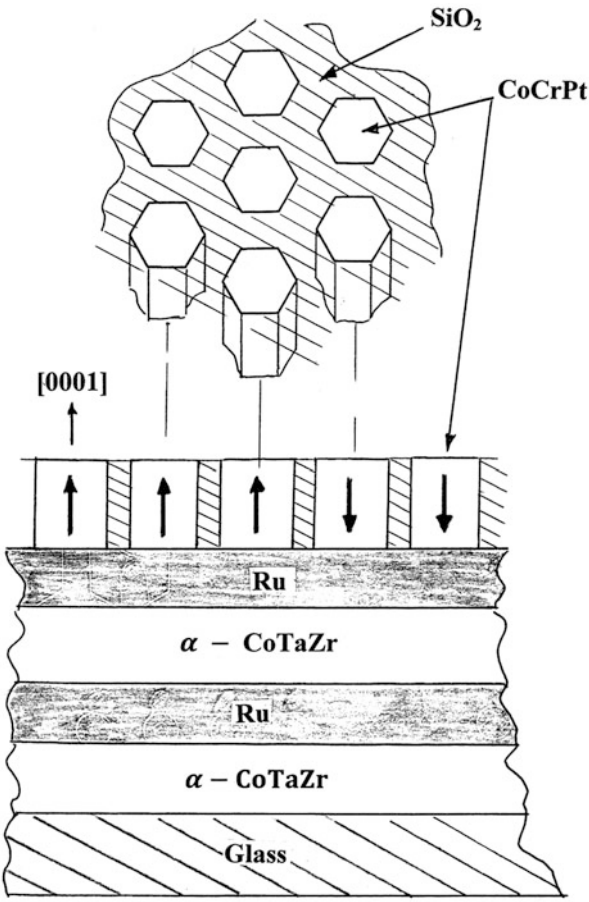


Fig. 22 Contemporary (2013) hard disk magnetic recording media composed of nanosize CoCrPt columnar [001] magnetic grains in SiO_2 matrix. Deposition starts with soft magnetic underlayer (CoTaZr) and Ru seed layers shown. [0001] texture gives rise to strong perpendicular magnetic anisotropy. Shown as ideal patterned media

are dispersed in SiO_2 and having a strong [0001] fiber texture as illustrated schematically in Fig. 22. These perpendicular bit arrays are deposited on a glass substrate with an amorphous CoTaZr soft magnetic underlayer and a Ru seed layer which optimizes the columnar grain size and [0001] orientation (easy axis of magnetization) of the CoCrPt grains (Fig. 22). Actually, one bit includes multiple CoCrPt grains magnetized in the same direction (0 or 1), usually around 10 grains for a good signal-to-noise ratio. To achieve a bit size of the areal density of ~1 T bit/in^2 (25 × 25 nm^2) will require ferromagnetic particles (or grains) of ~4–6 nm. This makes them thermally unstable because the magnetocrystalline energy becomes comparable to the thermal energy (kT). Overcoming this feature will require granular materials with a larger magnetocrystalline anisotropy such as cubic, ordered FePt (Varaprasad et al. 2013). Switching field intensity and speed of switching is also an important consideration. In this regard, smaller, magnetizable particles (grains) require higher magnetization (coercive) fields, and

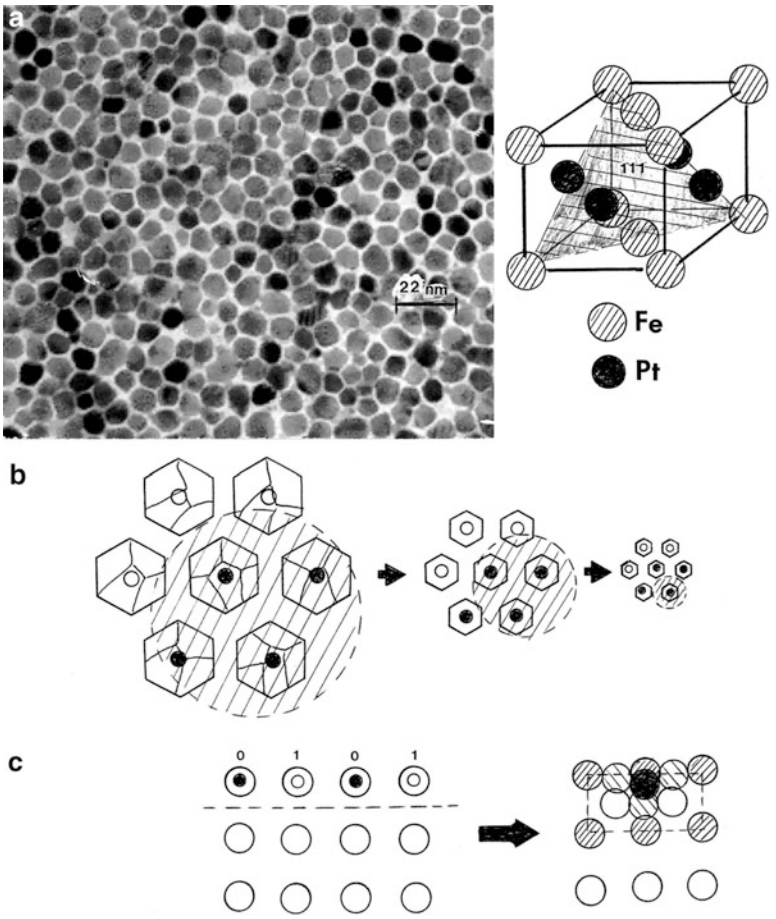


Fig. 23 Perpendicular magnetic recording media development. (**a**) Plane view TEM image showing FePt–C film grown on MgO underlayer on a thermally oxidized Si substrate. FePt and C are co-evaporated. FePt ordered fcc unit cell shown at right illustrates [111] texture of particles (Adapted from Varaprasad et al. 2013). (**b**) Hexagonal magnetic particles in hexagonal array converging to single nanocrystal particles as one bit recording (memory) area. Bit area is indicated by *shaded circles*. Dipole polarity is noted ● (down) or ○ (up): 0 or 1. Particles at *left* are polycrystalline (grain boundaries are indicated). (**c**) Square bit map or bit pattern with memory line word 0-1-0-1 converging to 12-atom bit volume in 3 layers: *open-shaded*-solid atoms. The bit pattern (*square*) at *left* consists of large bit areas incorporating millions of atoms polarized (magnetized) as shown (● : down; ○ : up)

magnetization switching must be assisted by some external energy. Heat is the most efficient external energy strategy which can be achieved using a focused laser beam. Coercivity is reduced as the temperature increases. This heat-assisted magnetic recording (HAMR) technology can be achieved by densely dispersed ferromagnetic particle regimes exhibiting high magnetocrystalline anisotropy as illustrated for an FePt–C granular thin film shown in Fig. 23a.

While Fig. 23a represents a relatively dense, granular magnetic medium, more regular, patterned or *bit-patterned media* for magnetic recording are also of interest (Tong and Lee 2010; Ranjbar 2012), as shown ideally in Figs. 22 and 23b. Figure 23b shows the convergence of the grain size and bit regime to a single grain for a hexagonal, close-packed array, while Fig. 23c shows this convergence for a square array with linear sequences of binary combinations or words. These arrays resemble the n × n magnetic core memory shown in Figs. 17b and 18a. In Fig. 23c, the bit volume is reduced to 12 atoms or a 12-atom pyramid-like quantum dot or spintronic dot where individual spins are oriented up or down, creating an atomic-scale magnetic memory as recently described by Loth et al. (2012), with the prospects for creating 150 TB drives.

Although magnetic core storage has not been used since the 1970s, magnetic tape in various forms is often used for large volume data storage because it is relatively cheap. Data storage, or length of storage, is also of some concern regarding the stability of the magnetic storage material and its magnetization. Heat and environmental conditions may alter the magnetization, especially where the magnetic material is subject to reaction or corrosion. Extreme magnetic environments may also destroy the magnetization as well.

Functional Magnetic Materials Innovations and Applications

Changes in electrical resistance in a material with an applied magnetic field magnetoresistance, giant magnetoresistance (GMR) or colossal magnetoresistance (CMR), magnetic field-induced isothermal entropy change in a magnetic solid-magnetocaloric effect (MCE), or giant magnetocaloric effect (GMCE) and magnetoelastic phenomena such as magnetostriction or ferromagnetic (magnetic) shape-memory effects (MSME), all contribute to a wide range of functional applications including sensors, transducers, actuators, and refrigerants. Applications at high temperatures are also important. GMR is applied in read–record heads illustrated schematically in Fig. 21b. Magnetic field control enables fast response in magnetic shape-memory alloys in contrast to more conventional shape-memory alloys discussed in chapter "► Implant Materials and Structures," Part X.

The most striking ferromagnetic shape-memory alloys are known as full *Heusler alloys* with the composition A_2BX, where A and B are transition metals, and X is a nonmagnetic element: Ni_2MnGa and Ni_2MnIn. These alloys have a cubic austenitic state and a different crystal structure in the martensitic state, the most prevalent being tetragonal. These alloys or other compositions of Heusler alloys such as $Ni_{50}Mn_{34}$ In_{16} exhibit large magnetocaloric effects as well as giant magnetoresistance, and these multifunctional properties can be shifted in temperature range by Cr doping in the Mn sites and Cu doping in the Ni sites (Sharma et al. 2010). Other room temperature magnetostrictive alloys include rare-earth Ni and Fe such as GdNi and GdFe. GdNi also exhibits a magnetocaloric effect, while GdSiGe and related rare-earth SiGe compounds exhibit giant magnetocaloric behavior near room temperature. Gd_5 $(Ge, Si)_4$ alloys exhibit GMCE and show a phenomenon

known as spontaneous generation of voltage (SGV) (Levin et al. 2001) which occurs in the vicinity of the first-order magneto-structural phase transition. These multifunctional materials are applicable to sensors of temperature, pressure, and magnetic fields. Similar MCE and GMCE behavior is also observed for La (Fe, (Si, Ge))$_{13}$. In thin film form, these alloys can have applications in thin film refrigeration (cooling) for electronic device applications, and biocompatible magnetocaloric materials can play a role in medical applications such as hypothermia (Tishin and Spichkin 2003). Other MCE materials include ABO_3 (perovskite) magneto-oxides such as $(La_{0.5}Ca_{0.5})$ MnO_3, $(Pr_{0.5}Sr_{0.5})$ MnO_3, $(Nd_{0.5}Sr_{0.5})$ MnO_3, and $LaCoO_3$, which exhibit high-temperature magnetism unlike other MCE materials, especially Sr-doped $LaCoO_3$ (Medling et al. 2012). Magnetic oxide materials are also used in a wide range of high switching frequency (>0.5 GHz) applications due in part to their high resistivities.

While rare-earth (RE)-related alloys and compounds represent many prominent multifunctional magnetic materials, compounds such as Fe_{16} N_2 have been shown to rival their superior magnetic properties. Furthermore, the rare-earth market is a volatile one, and it has become expedient not only to look at RE-materials substitutes but also to reexamine prospects for novel processing of more traditional magnetic materials such as Alnico alloys to achieve superior magnetic behavior (Palasyuk et al. 2013).

References

Baibach MN, Broto JM, Fert A, VanDau FN, Petroff F, Eitenne P, Gruzet G, Fredrich A, Chazelas J (1988) Giant magnetoresistance of (001) Fe/(001)Cr magnetic superlattices. Phys Rev Lett 61(21):2472–2475

Cardarelli F (2005) Encyclopedia of scientific units, weights, and measures, their SI equivalences and origins. Springer, Heidelberg

Daniel ED, Mee CD, Clark MH (1998) Magnetic recording, the first 100 years. Wiley, New York

Engdahl EG (2000) Handbook of giant magnetostrictive materials. Academic, San Diego

Forrester JW (1951) Digital information in three-dimensions using magnetic cores. J Appl Phys 22:44–48

Grünberg P, Schreiber R, Pang Y, Brodsky MB, Sowers H (1986) Layered magnetic structures: evidence for antiferromagnetic coupling of Fe layers across Cr interlayers. Phys Rev Lett 57:2442–2445

Johnson M, Silsbee RH (1985) Interfacial charge-spin coupling: Injection and detection of spin magnetization in metals. Phys Rev Lett 55(17):1790–1793

Khizroev S, Litvinov D (2005) Perpendicular magnetic recording. Kluwer Academic Publications (Springer), Dordrecht

Levin EM, Pecharsky VK, Gschneider KA (2001) Spontaneous generation of voltage in Gd_5 (Si_xGe_{4-x}) during first-order phase transition induced by temperature or magnetic field. Phys Rev B 63(17):174110–174117

Loth S, Baumann S, Lutz CP, Eigler DM, Heinrich AJ (2012) Science 335(6065):196–199

McHenry ME, Willard MA, Laughlin DE (1999) Amorphous and nanocrystalline materials for applications as soft magnets. Prog Mater Sci 44:291–433

Medling S, Lee Y, Zheng H, Mitchell JF, Freeland JW, Harmon BN, Bridges F (2012) Evolution of magnetic oxygen states in Sr-doped $LaCoO_3$. Phys Rev Lett 109(15):15704–15709

Mee CD, Daniel ED (eds) (1996) Magnetic recording technology. McGraw-Hill, New York

Meyer CD, Bedair SS, Morgan BC, Arnold DP (2010) High-inductance density, air-core, power inductions and transformers designed for operation at 100–500 MHz. IEEE Trans Magn 46(6):2249–2254

Murr LE (1978) Solid-state electronics. Marcel Dekker, New York

Palasyuk A, Blomberg E, Protorov R, Yue L, Kramer MJ, McCallum RW, Anderson IE, Constantinides S (2013) Advances in characterization of non-rare earth permanent magnets: exploring commercial Alnico grades 5–7, and 9. JOM 65(7):862–869

Pecharsky VK, Gschneidner KA Jr (1997) Giant magnetocaloric effect in Gd_s (Si_2Ge_2). Phys Rev Lett 78(23):4494–4497

Phan M-H, Peng H-X (2008) Giant magnetoimpedance materials: fundamentals and applications. Prog Mater Sci 53:323–420

Ranjbar M (2012) Investigation of magnetic nanostructures for patterned media. Lambert Acad. Publishers, Saarbrücken

Reilly ED (2003) Milestones in computer science and information technology. Greenwood Press, Wesport

Sharma VK, Chattopadhyay MK, Khendelwal A, Roy SB (2010) Martensitic transition near room temperature and the temperature-and-magnetic-field-induced multifunctional properties of $Ni_{49}CuMn_{34}In_{16}$ alloy. Phys Rev B 82(17):172411–172414

Shinjo T (2009) Magnetism and spintronics. Elsevier, New York

Tishin AM, Spichkin YI (2003) The magnetocaloric effect and its applications. IOP Publishing, Philadelphia

Tokura Y (2000) Colossal magnetoresistive oxides. Gordon and Breach Science, New York

Tong DD, Lee Y-J (2010) Magnetic memory: fundamentals and technology. Cambridge University Press, Cambridge, UK

Varaprasad BSDCS, Takahashi YK, Hano K (2013) Microstructure control of LI_0-ordered FePt granular film for heat-assisted magnetic recording (HAMR) media. JOM 65(7):853–861

Wolf SA, Chtchelkanova AY, Treger DM (2006) Spintronics – a retrospective and perspective. IBM J Res Dev 50:101–112

Yunas J, Majlis BY (2008) Comparative study of stack interwinding micro-transformers on silicon monolithic. Microelectron J 39(12):1564–1567

Yunas J, Homzah AA, Majlis BY (2009) Surface micromachined on-chip transformer fabricated on glass substrate. Microsyst Technol 15(4):547–552

Zhou Y, Yu JQ, Zhao XL, Cai BC (2001) Giant magnetoimpedance in layered FeSiB/Cu/FeSiB films. J Appl Phys 89:1816–1819

Electronic and Optoelectronic Materials and Device Innovations

Contents

Abstract

This chapter begins with solid-state drives and flash memory materials and device structures. Thin-film transistor (TFT) innovations are described. TFT-LCD are discussed along with photodiode and photovoltaic materials and structures, including LED display concepts. Photovoltaic materials and devices – solar cells and multijunction solar cell arrays – are described along with dye-sensitized solar cells. Thermionic materials and devices are described along with innovations in hybrid photovoltaic/thermoelectric systems and devices. Battery technologies and energy storage materials, including thin-film battery concepts which include Li-ion and Li–air batteries, are described. Supercapacitors and supercapacitor energy storage materials and device innovations are discussed followed by sensor materials and device concepts and fundamentals.

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_58

Introduction

While there are continuous streams of electronic and optoelectronic materials and device innovations, there are some notable and commercially visible innovations which will be highlighted in this chapter. Consumer electronic products involving computational tools, including of course solid-state electronic memory of various kinds as well as display technologies ranging from LED to OLED arrays measuring a few centimeters to many meters, are among the more obvious. These are followed by energy-related technologies involving photovoltaic and thermoelectric materials and devices, including hybrid photovoltaic/thermoelectric energy conversion systems utilizing innovations in inorganic and organic materials or composites, ranging from thin-film arrays to large-area solar collection power stations. Supercapacitors and battery materials technologies, including thin-film power storage for electronic devices and a variety of electric vehicles, also pose novel materials and materials systems innovations.

The integration of complex and multifunctional materials and materials systems into a variety of electronic device structures and energy systems such as electric automobile or automotive hybrids along with a variety of aeronautical and biomedical systems and materials structures has encouraged innovations in sensors and sensor materials for a wide range of interactive diagnostics. These include variations of thin-film arrays including metal oxides and other materials whose microstructures, properties, and processing interrelationships can be manipulated to produce performance functionalities involving chemical, thermal (temperature), mechanical, radiation, moisture, and related diagnostic sensing based on the systematic manipulation of resistive, capacitive, inductive, and other fundamental surface properties of materials.

Materials Aspects of Solid-State Drives (SSDs): Flash Memory

Since magnetic memory concepts are only a few pages back in chapter "► Innovations in Magnetic Materials," it would seem expedient to discuss materials aspects and design principles of solid-state memory or drives. Indeed, solid-state or transistor-based memory is not a drive as implicit in magnetic hard disk drives discussed in chapter "► Innovations in Magnetic Materials." SSD technology, and especially flash memory, uses no moving mechanical components as in hard disk drives (HDDs) (Micheloni et al. 2013). Most contemporary SSDs use NAND-based flash memory in the form of memory cards illustrated in Fig. 19 (Micheloni et al. 2010). However, contemporary flash memory capacities exceed 240 MB in contrast to 8 MB shown in Fig. 19 overlain on the original magnetic core memory array structure used in early computer systems. NAND-based (or NAND logic-based) flash memory uses random access memory (RAM) with separate power sources to maintain embedded, nonvolatile memory. NAND logic refers to inverted AND logic. Correspondingly, NOR logic refers to inverted OR logic, with AND-OR referring to storage in transistor arrays or bit arrays as

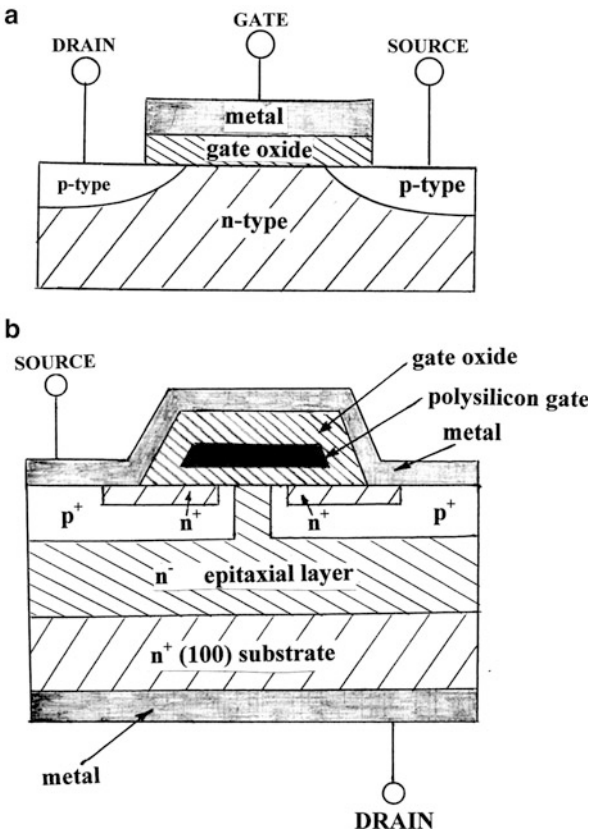


Fig. 1 Comparison of standard MOSFET structure (**a**) with floating gate NAND flash memory (FET) cell (**b**). Note material/layer zones. Also note that p^+ and n^+ denote heavy or high doping (for *p*-type and *n*-type semiconductor regions), while n^- denotes especially low *n*-type doping levels

0 or 1. Flash memory developed over the past roughly 4 years (since 2010) from electrically erasable programmable read-only memory (EEPROM). NAND-type flash memory may be written and read in so-called blocks or pages much smaller than the device. Correspondingly NOR-type flash memory devices allow a single machine word (or byte) to be written or read independently.

In flash memory, each memory cell (or bit) resembles a standard MOSFET illustrated in Fig. 1, where Fig. 1b shows a discrete MOSFET as an integrated bit zone in an IC. In the device structure shown in Fig. 1b, a polysilicon layer surrounded or sandwiched by oxide (SiO_2) serves as a single gate as shown in the TEM cross-sectional view in Fig. 2. Flash memory transistor cells are fabricated with two gates: a control gate as in standard MOS transistors and a floating gate below it, isolated by insulating oxides and composed of polysilicon (Fig. 2) as illustrated schematically in Fig. 3a. The ONO dielectric layer in Fig. 3a is oxide–nitride–oxide composition, SiO_xN_y or $SiO_2/Si_3N_4/SiO_2$, which is an effective electron trapping of direct tunneling of electrons and holes from the poly-Si for programming (storage) and erasing of data. When charge is applied to or removed

Fig. 2 Cross-sectional TEM (XTEM) view of polysilicon layer sandwiched between two amorphous SiO_2 insulating layers in device structure as in Fig. 1b. Contemporary, MOSFET cells have poly-Si layers roughly one-tenth the thickness

from the floating gate, the transistor threshold voltage changes as shown in Fig. 3b and c respectively, and it is this voltage change which embodies memory. As shown in Fig. 3b, c, to program or erase a flash memory cell requires a set of pulses to be applied in a time-controlled manner. Injected charge (Fig. 3b) or removed charge (Fig. 3c) from the floating gate creates a memory state (0 or 1) which remains even after power is removed, a nonvolatile memory.

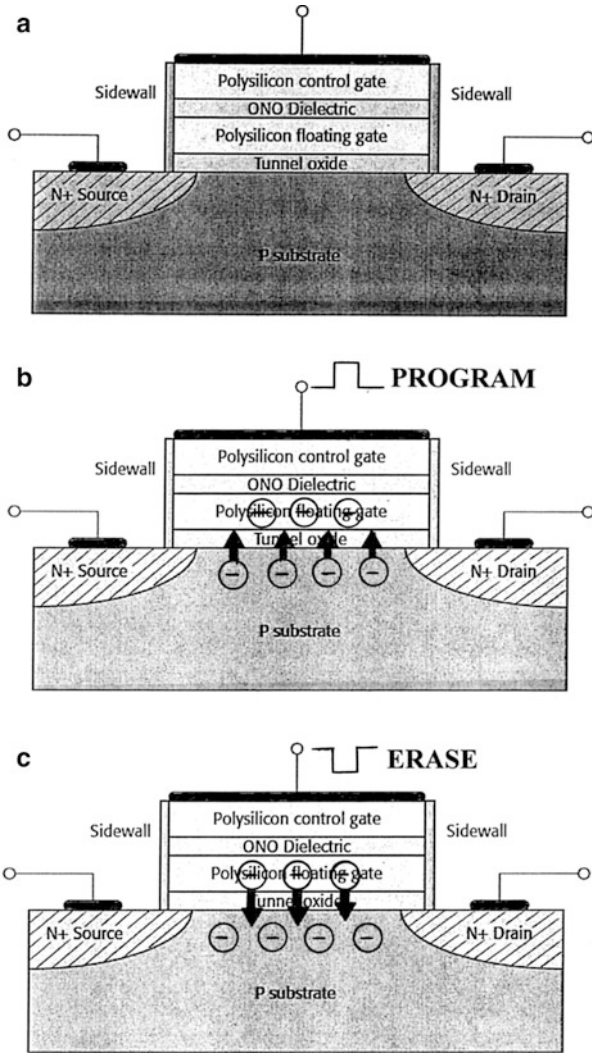


Fig. 3 Floating gate, NAND flash memory cells showing programming (setting) and resetting or erasing memory elements. (**a**) Basic floating gate NAND cell as in Fig. 1b. (**b**) Charge transfer to set or program cell. (**c**) Charge transfer to reset or erase cell

Figure 4 compares NAND and NOR flash cell array features of flash memory. The NAND cell is reduced to the NOR cell design where a typical NAND array of MOSFETs has a unit block typically formed by 16 NAND cells in series with two select transistors: the ground select transistor and the bit line select transistor. The NAND architecture has high bit densities but slow read times and is therefore more appropriate for flash memory storage applications. NAND flash is similar to hard disk drive and NAND flash cells (Fig. 4) are 60 % smaller than NOR cells because NOR flash cells require a separate contact for each cell. Pikus and Likharev (1997) has described a nanoscale field-effect transistor (FET) to have a MOSFET minimum size of ~4 nm. This approximates the perpendicular magnetic memory bit array size illustrated in Fig. 23a, for example.

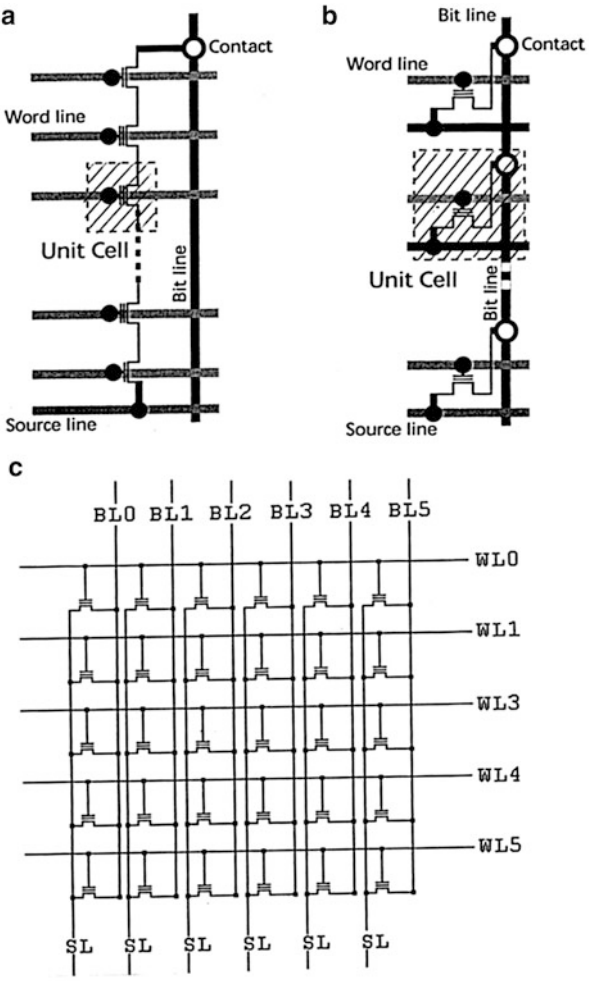


Fig. 4 Comparison of NAND (**a**) and NOR (**b**) flash (memory) cells. Note contacts along bit line. Note bit and word addressing modes are possible with either NAND or NOR flash cells. (**c**) NOR array layout section showing source lines (*SL*), bit lines (*BL*), and word lines (*WL*). A single word line is asserted to read a cell. The source lines are asserted and the assertion or deassertion of a bit line gives the contents of the storage cell line the same row. Turning a circuit's current flow on and off allows individual bits to be routed to a data bus

There are specific advantages and disadvantages to using NAND or NOR flash memory devices in embedded systems. NAND is more suitable for sequential or file data, while NOR is better suited for random access and byte-write capabilities. NAND flash is used in virtually all removable cards: USB (or thumb) drives, secure digital (SD) cards, memory stick cards, and multimedia cards (MMCs). NAND blocks or flash blocks will normally survive $\sim 10^5$ program/erase cycles. Flash memory is temperature and shock resilient, but can be compromised by high-energy radiation which can cause defects in the transistor (cell) crystal structures and alter the charge or bit resolution and charge tunneling efficiency.

Thin-Film Transistor (TFT) Innovations

It can be observed in Figs. 1, 2, and 3 that all transistors, solid-state device, or ultra large-scale integrated (ULSI) circuit transistor components are thin-film (multilayer) structures. All are also field-effect transistors (FETs). Also as illustrated above (Figs. 1, 2, 3, and 4), the principle of contemporary MOSFET-based memory technologies is gate control of current flowing from the source to the drain (Fig. 1a). In the case of a floating gate or a second gate complementing the control gate (Fig. 3), which is connected to the bit line (Fig. 4), the floating gate in fact "floats" since it is not connected to anything. However when a bit is programmed, electrons are stored upon the floating gate, and the floating gate is able to store a charge of its own. This adds to the control gate bias, and if there is a change on the floating gate, then no current will flow through the channel between source and drain whether or not a charge exists on the control gate. Of course normal transistors as illustrated in Fig. 1a without a floating gate can act as conductivity control devices rather than logic devices. Since the transistor, and especially as a memory bit, is discrete in terms of the space it occupies in the chip landscape or logic array as shown schematically in Fig. 4c, the memory density is also discrete in contrast to magnetic hard drive memory where it is not possible to activate discrete, columnar grains because the bit size is a grouping of oriented dipole volumes.

Although solid-state IC devices can be mounted on large and even flexible areas, thin-film transistors along with other components having areas larger than conventional, lithographed dies, and especially organic thin-film transistors (OTFTs) can be printed with ink jet technologies as described in chapter "► 3D Printing: Printed Electronics." OTFT innovations are not only able to create memory cells using a floating gate concept, but they can also produce other optoelectronic configurations, including organic LEDs (OLEDs) and related pixel access devices in active matrix displays. These applications use organic semiconductors as the active material or active channel. Currently, pentacene ($C_{22}H_{14}$), a polycyclic aromatic hydrocarbon containing five linear benzene rings, exhibits high-performance features in the two common OTFT-FET structures shown conceptually in Fig. 5a, b. Figure 5c shows the molecular fragment structure for the highly conjugated pentacene along with recent AFM images (Gross et al. 2009). Figure 5d shows a Si-bottom gate OTFT (or OFET) schematic with a photosensitive polyimide modification layer inserted between the gate dielectric and pentacene layer which manipulates charge transport, similar in some respects to the floating gate shown in Fig. 3.

While TFT structures illustrated in Fig. 5 can be accommodated on large and moderately flexible surfaces, truly flexible polymer-base formats or even paper will require substitution of silicon-base configurations by suitable semiconducting polymers as shown, for example, in OLED configurations in chapter "► 3D Printing: Printed Electronics." Figure 6 illustrates a conceptual modification for Fig. 5a, b, d with polymeric films replacing the SiO_2 dielectric and Si base and with graphene serving as the source–drain–gate electrodes. Graphene has high conductivity, higher than copper even as a single layer, and can be applied by spin coating or CVD.

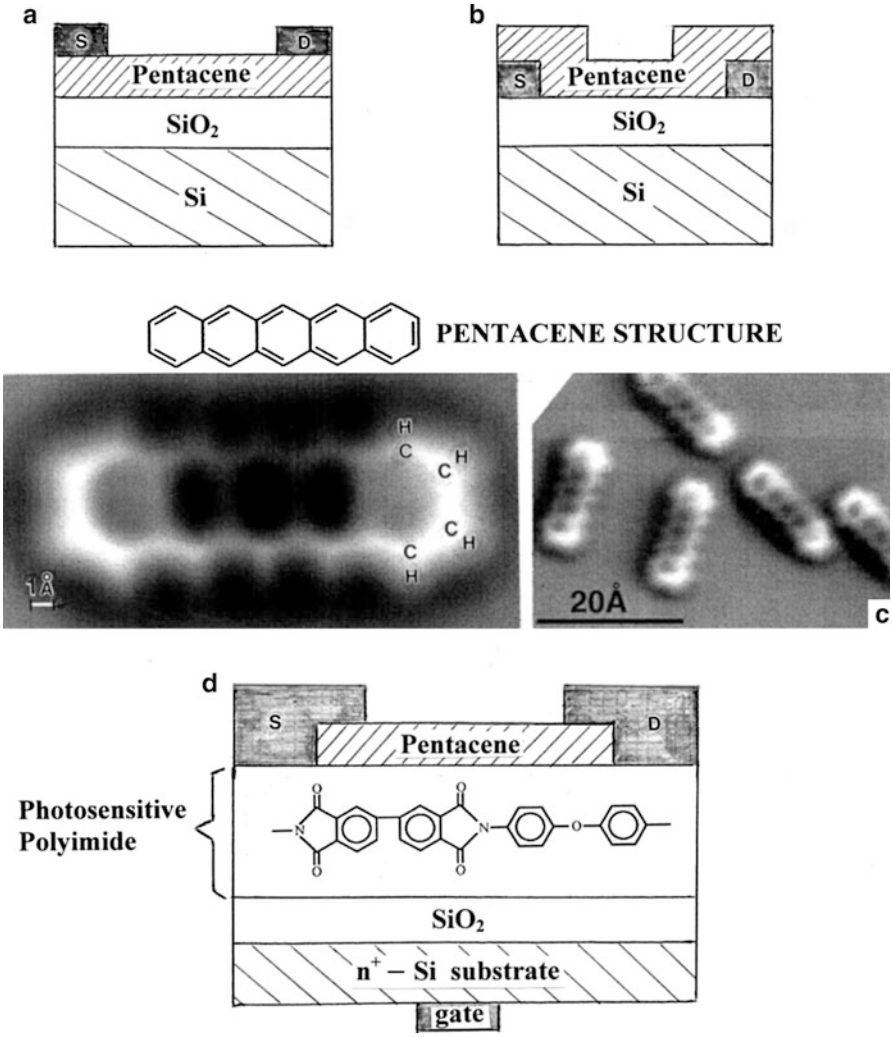


Fig. 5 Pentacene-based TFTs. (**a**) and (**b**) Show *top* and *bottom* contact structures, respectively. S and D are source and drain. (**c**) Shows the pentacene structure and AFM images adapted from Gross et al. (2009). (**d**) Shows pentacene TFT with photosensitive polyimide modification layer with structure shown in insert. The n^+-Si substrate acts as the gate electrode

It can also provide contact compatibility with the other polymeric components and act as a contact for metal transitions as well because polymer/metal contacts are often difficult to achieve because of metal/polymer interfacial phenomena. In addition, since a graphene monolayer has an optical transparency of ~98 % (Nair et al. 2008), it can replace the more common transparent, conducting SnO_2 or indium tin oxide (ITO) films used in device windows such as photovoltaic (solar cell) structures or LED structures illustrated in chapter "3D Printing: Printed Electronics"

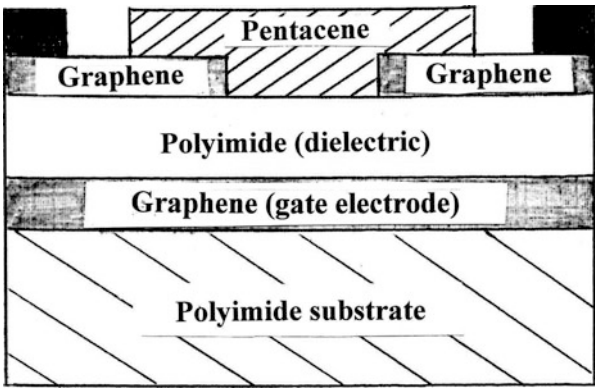


Fig. 6 Flexible OFET structure using source–drain–gate graphene electrodes and only polymeric TFT components. Note the general polyimide structure is shown in Fig. 5d

(see Fig. 4). These structures will be explored in more detail later in this chapter as well. Li et al. (2013) have recently discussed the integration of graphene into TFTs, including OFETs as illustrated in Fig. 6, and other graphene FET arrangements for solid-state structures are also being explored. In addition, MoO_3 nanosheets can also compete with graphene in the context of electron flow, while hexagonal BN in nanosheet form is a good insulator. Consequently nanosheet structures may be processed as nano-TFTs for some applications.

Optoelectronic Materials and Device Innovations

Optoelectronic devices can be broadly described as homojunction (and sometimes heterojunction) converters (transducers) or controllers of light: optical to electrical or electrical to light, where light can include a broad spectrum of wavelengths, both visible and invisible (X-rays to infrared). Figure 7 illustrates the more fundamental photoconduction and photoconversion device concepts, beginning with simple photoresistors such as CdS (Fig. 7a) and ending with the solar cell (or photovoltaic) device principle. Figure 7b, c shows light directed to the device (optical to electrical) or from the device (electrical to optical). Figure 7b illustrates the classical junction (n-p) diode representing the photodiode (optical to electrical) or LED (electrical to optical). The n-p or p-n junction device can also function as a junction laser illustrated schematically in Fig. 8a, while Fig. 8b illustrates a double heterostructure which is also edge emitting and can function as an edge-emitting LED as well. In contrast, Fig. 8c illustrates a surface-emitting LED heterojunction structure. Laser diodes are fundamentally similar to LEDs as illustrated in Fig. 8, since both have an active region where electrons and holes recombine to emit light (Fig. 7d). However, a working laser diode also contains an optical cavity or waveguide structure where stimulated emission takes place. This waveguide is terminated by a mirror at each end to reflect photons back and forth as they are amplified by stimulated emission: the photon triggers radiative recombination of an electron and hole creating an additional photon with the same phase and energy as

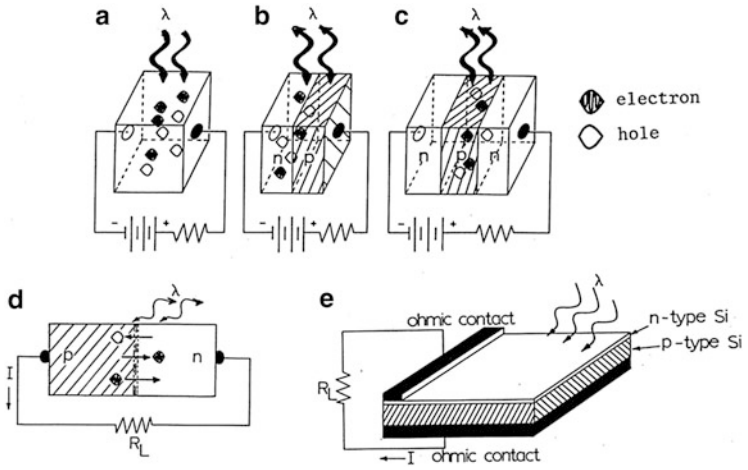


Fig. 7 Optoelectronic device examples. (**a**) CdS photoresistor (or photoconductor). (**b**) Photo (n-p) diode or LED. (**c**) Phototransistor. (**d**) Photovoltaic current generation at *p-n* diode junction. (**e**) Simple doped Si photovoltaic wafer (Murr 1978)

the incident photon, forming a coherent beam with wavelength λ. Solid-state lasers utilizing these simple principles have been readily available for the past two decades, often requiring only small battery power in handheld devices such as laser pointers.

LED technology, and especially display technology, is prolific, from small cm-sized flat or flexible display formats to huge multimeter displays. LED lighting is also becoming a viable substitute for fluorescent lighting as well. LED display arrays and their material components (inorganic, solid-state, and organic OLEDs) are simpler and more reliable than LCD or TFT-LCD array technology as illustrated on comparing an LED display array in Fig. 9 with a TFT-LCD flat-panel display array segment shown in Fig. 10.

Both Figs. 9 and 10 illustrate red–green–blue (RGB) matrix or pixel arrays. In Fig. 9, inorganic (or solid-state) LED units consist of addressable red (AlGaInP), green (InGaN), or blue (InGa(N)) diodes in an 8 × 8 unit (pixel) array (8 rows and 8 columns). Activation of all three diodes in the pixel unit produces white light. While LEDs can be manufactured in a wide range of dimensions, OLED technology has advantages in cost and production using 3D printing technologies described in detail in chapter "► 3D Printing: Printed Electronics," especially in portable device displays such as mobile phones and cameras and high-resolution (or high-definition) television production. Figure 11 illustrates large LED (inorganic) display technology in contrast to OLED television display resolution (Chen et al. 2012).

The most common liquid-crystal display technologies employ a thin-film transistor liquid-crystal display (TFT-LCD) consisting of a sandwich-like structure

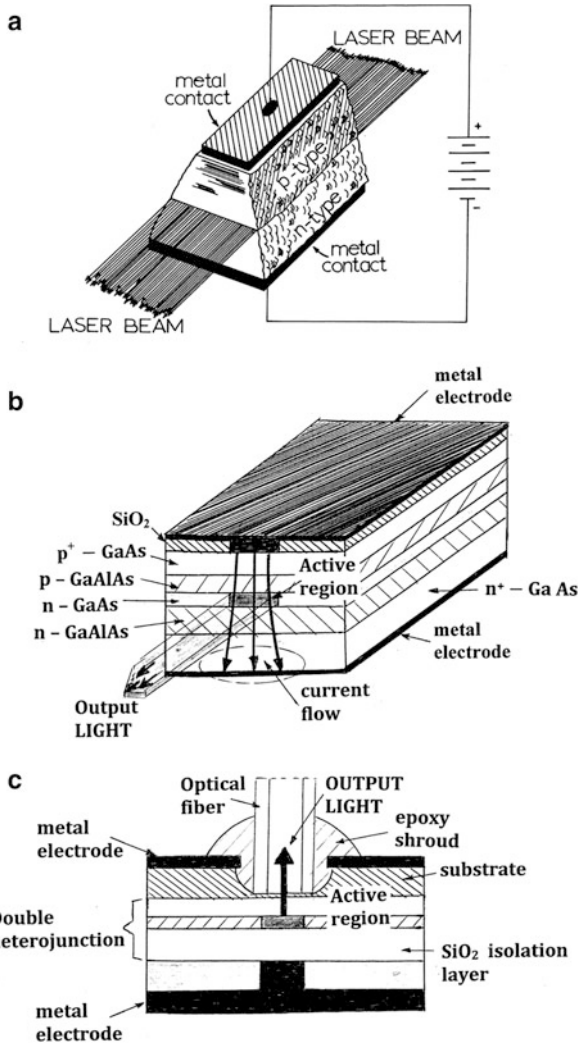


Fig. 8 Junction photoemitters. (**a**) Edge-emitting photodiode (laser or LED). (**b**) Edge-emitting heterostructure LED. (**c**) Surface-emitting LED or laser double heterojunction. See Fig. 1 caption for n^+, p^+ meaning

with liquid crystal filled between two glass plates (Castellano 2005). TFT-LCD technology is more complex and somewhat limited in contrast to LED or OLED display technology, usually limited to smaller display applications, with television screens being the largest. LED systems are bright enough so they do not require backlighting necessary for LCD technologies.

A TFT-LCD panel contains a specific number of subpixels or unit pixels having a TFT, a pixel electrode, and a storage capacitor (Cs) (Fig. 10). As shown in Fig. 10, a yellow pixel is lit by polarizing red and blue subpixels at the bottom of the 3×3 pixel array. Figure 12a shows a view of the unit pixel sandwich structure along with its equivalent circuit corresponding to the array of Fig. 10. Figure 12b shows additional details of the subpixel structure and the liquid-crystal on–off state.

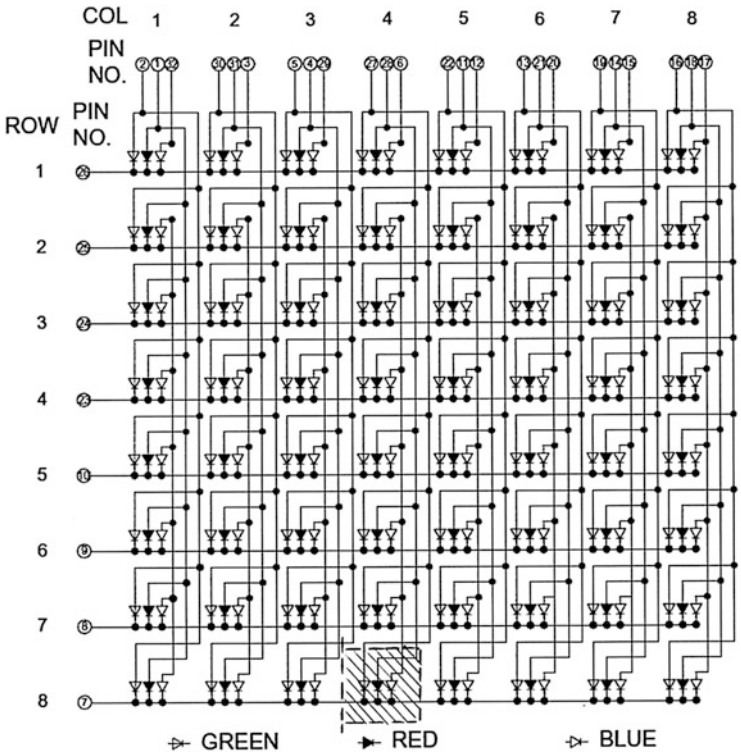


Fig. 9 Red–green–blue LED common cathode 8 × 8 pixel matrix display section circuit diagram. Pixel (*color dot*) is shown shaded. Colors other than R, G, and B can be generated by turning on combinations of LEDs. For example, R + B = Y (yellow). R + G + B = white (Adapted from Futurlec (electronic components superstore))

The elements of the color filter line up with the unit pixels. Each pixel in a color LCD contains three unit or subpixels corresponding to the RGB LED sequence illustrated in Fig. 9, forming color dot arrays. Any color, with some qualifications, can be produced by mixing these primary colors.

The most common semiconducting (TFT) layer as shown in Fig. 12a is made of amorphous Si although poly-Si has also been used. The color filter shown in Fig. 12a, b contains a black matrix and resin film containing red, green, and blue dyes or pigments. The liquid-crystal (LC) layer is injected between the transistor backplane and the color filter array. A transparent (indium tin oxide (ITO)) electrode is placed on the back of the front polarizer and behind the LC layer.

Common LCD or TFT-LCD as shown in Figs. 10 and 12 are referred to as a twisted nematic LC device where the surface alignment directions at the two ITO electrodes (Fig. 12) are perpendicular to one another, and the LC molecules arrange in a helical or twisted structure as shown in Fig. 12a, b. The polarizing filters are parallel and perpendicular, and without the liquid-crystal layer between them, light passing through the first filter would be blocked by the second crossed polarizer.

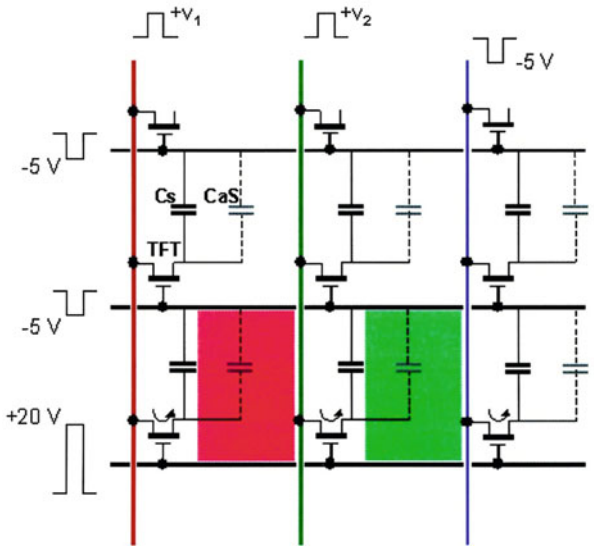


Fig. 10 Schematic view of 3 × 3 color TFT-LCD cell array. A yellow pixel is lit by polarizing the *red* and *blue* subpixels as shown. Horizontal lines drive the TFT (FET) gates, while vertical lines generate gray levels through the source contacts. Cs is the subpixel storage capacitance, while Cgs represents the parasitic capacitance between the cell electrode and front electrode, which are transparent, ITO-conducting films. Applicable voltages are illustrated along vertical and horizontal drives (Courtesy of Wikimedia Commons)

The twisted LC layer induces the rotation of the polarization of incident light and the display appears gray. Applying a sufficient voltage can untwist the molecules, and the polarization of the incident light is not rotated as it passes through the LC layer, and this light will be mainly polarized perpendicular to the second filter and blocked, causing the pixel to appear black. Correspondingly, controlling the electrode voltage in each pixel can allow light to pass through in varying intensities to create different gray levels. Since LCD panels produce no light of their own, they require external or back light as shown in Fig. 12b. Backlights are usually small fluorescent tubes inside the lower portion of the display. The LC molecules simply provide on–off or a range of on providing intensity variation through voltage control noted above.

Liquid crystals were discovered by Friedrich Reinitzer, an Austrian botanical physiologist, in 1888 in studies of various derivatives of cholesterol, later referred to as cholesteric liquid crystals. Especially notable was the observation of two melting points in cholesteryl benzoate. Later observations noted two other important features of cholesteric liquid crystals: the reflection of circularly polarized light and the ability to rotate the polarization direction of light. Liquid crystals can be divided into three phases: thermotropic, lyotropic, and metallotropic. Thermotropic and lyotropic phases consist of organic molecules, while metallotropic phases are composed of both organic and inorganic molecules, and the liquid-crystal transition depends not only on temperature and concentration but also on the organic–inorganic composition ratio (Chandrasekhar 1992). At low enough temperatures, most liquid crystals will form a conventional crystal. Consequently, they are a molecular array which lies between a liquid (no order) and a solid crystal (regular order, which can be short range or long range). The term mesophase is used to describe subphases of liquid crystals: nematic, smectic, chiral, blue, and discotic phases.

Fig. 11 LED displays. (**a**) Large building flexible (*outdoor billboard*) display screen in China (Courtesy Best-LED display, China). (**b**) OLED television screen (Courtesy of LG, South Korea)

The distinguishing feature of liquid-crystal molecules is their rodlike, rigid, dipolar structure. Nematic phase LCs which are characteristic of liquid-crystal materials used in LCD or TFT-LCD technologies shown in Fig. 12 are typical of this structure feature and easily aligned by both an external electric and magnetic field (Chandrasekhar 1992; Castellano 2005).

It might be observed in Fig. 12b that TFT-LCD fabrication in flexible display systems might be facilitated by replacing layers with flexible polymers, including OFET structures and graphene electrodes (Fig. 6). However, as noted previously, OLED fabricated displays (Fig. 11b) are more cost-effective and can accommodate large and flexible display requirements.

It is interesting to note that IC logic or chip arrays, memory arrays (Fig. 4), and pixel arrays (Figs. 9 and 10) are all characterized by some form of Moore's Law (Brock 2006): the nearly doubling of areal densities in 1–2 years and the corresponding reduction in cost. This is notable in microprocessor transistors/chip, magnetic areal storage (Gb/cm^2), display and digital camera resolution (GB or pixels/cm^2), etc., and bits/$, pixels/$, $/transistor, etc. At some point in

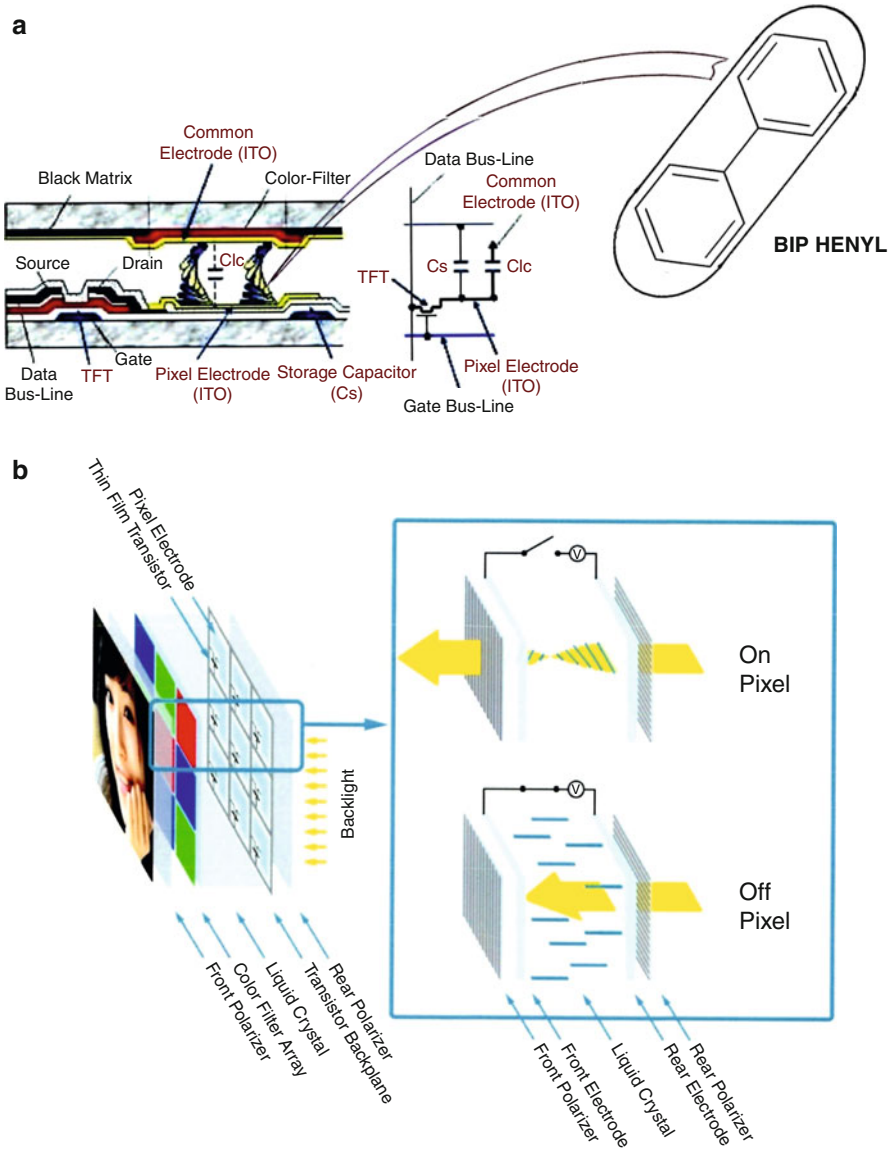


Fig. 12 TFT-LCD structure schematics. (**a**) Unit pixel structure view and equivalent circuit, storage capacitor (Cs), and equivalent liquid-crystal capacitor (Clc) are connected as load. Applying a positive pulse to a gate electrode turns TFT on, charging Cs and Clc and twisting liquid-crystal molecules. Common LC biphenyl molecular structure is shown at *right*. The hexagonal rings can also rotate 45° relative to one another relative to one another. (**b**) Details of pixel on-off feature and sandwich structure components (Courtesy of Applied Materials, Austin, TX)

time, as predicted at least over the past decade, the slope of such linear or log-linear trends will change just as the metrics shift over time to accommodate new innovations and related technologies. It is certainly apparent that new materials and materials innovations will continue to drive these technologies and their associated economic impacts. Functional size regimes such as atomic bit volume illustrated in Fig. 23c or the development of single molecule organic electronics or devices may provide a point of convergence, after which a new paradigm for development and innovation may emerge (Heath 2009).

Photovoltaic Materials and Device Innovations

The shining of light on junction (n-p) arrays as illustrated in Fig. 7 can generate electrons and holes which can create an appropriate current flow if recombination can be suppressed or avoided. Exposing large surface areas to sunlight as illustrated schematically in Fig. 7e can result in increasing conversion efficiency to electricity. While doped (n or p) Si in various forms has been a common photovoltaic material, other junction compounds and doped compounds can be used in creating photovoltaic or solar cell devices along with a wide range of organic compounds to create organic photovoltaic cells (OPVCs) which often have the advantage of being spreadable and correspondingly simpler to manufacture, especially in flexible configurations. However, the efficiency of solar conversion for organic solar cells is generally half or less than corresponding inorganic, solid-state solar cells. Figure 13 illustrates the general concept for a photovoltaic (solar) cell.

The theoretical efficiency limit of 33.5 % of a solar cell was initially determined by Shockley and Queisser (1961) in the context of the bandgap (or a single bandgap), and they established a balance equation between photons entering and escaping the semiconductor and the electron–hole pairs produced and extracted as electrical current. More recently Landsberg and Torge (1980) determined the limiting efficiency for solar conversion to be given by

$$\eta(\text{Landsberg}) = 1 - \frac{4}{3}\left(\frac{T_{\text{ambient}}}{T_{\text{sun}}}\right) + \frac{1}{3}\left(\frac{T_{\text{ambient}}}{T_{\text{sun}}}\right)^4 \quad (1)$$

where the ratio $T_{ambient}/T_{sun}$ = 300 K/600 K or 0.05; and these temperatures are derived from considering that a solar energy converter is connected to the sun which is a reservoir of photons at ~600 K, while the converter is connected to a cold reservoir, the Earth, at ~300 K. With this ratio, Eq. 1 gives an efficiency of 93.3 % in contrast to the Carnot efficiency of 95 %. However, more recently Araujo and Marti (1994) calculated a top efficiency of 40.7 % for blackbody radiation at 600 K/300 K (solar/ambient) temperatures. As illustrated in Fig. 14, efficiencies above 40 % have been recorded up to 2010 for experimental solar cells, and there are currently a great variety of commercial solar cell arrays or large-scale solar-photovoltaic panel arrays as illustrated in Fig. 15. These achieve efficiencies in the 10–20 % range.

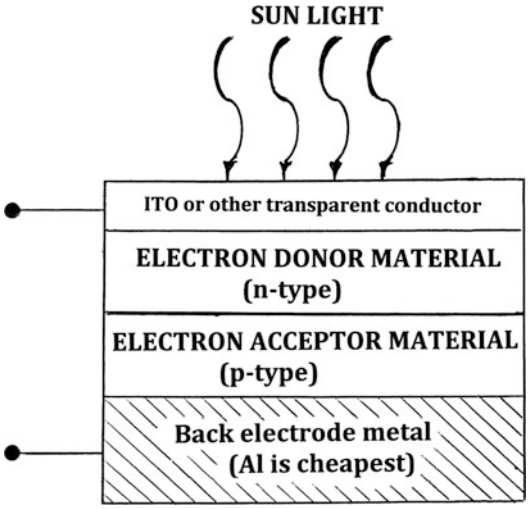


Fig. 13 Simple photovoltaic (solar) cell concept. The transparent (conducting) electrode is often indium tin oxide (ITO) or comparable material. The electron donor (*n*-type) or hole acceptor (*p*-type) material can be inorganic (solid-state) material or organic material in so-called bulk (layer) or thin-film form

Antonio Luque (2011) has recently intimated that the reason solar cells operate so far from the Landsberg efficiency (η (Landsberg)) expressed in Eq. 1 is because they do not use the entire solar spectrum, only photons with energy above the bandgap. In addition, entropy is generated in cooling of the electron–hole pairs produced by high-energy photons; in fact they are delivered at the cell voltage which cannot exceed the bandgap energy. In the ideal solar cell, the open circuit voltage (V_{oc}) is given by

$$V_{oc} = E_g(1 - T_{ambient}/T_{sun}) \quad (2)$$

where E_g is the bandgap energy. But no power is extracted for a solar cell in open circuit. To achieve zero or near zero entropy, an ideal photovoltaic cell could be covered by a narrow band-pass filter and illuminated by monochromatic radiation.

A novel concept for overcoming these bandgap efficiency issues is to stack solar cells of different bandgaps in order to increase the solar conversion spectrum or in fact utilize the entire solar spectrum for electricity generation. Additionally, the upper cells of such stacked arrays can act as a filter for the lower cells, where each cell is illuminated by a narrow range of photon energies, thereby also reducing the entropy production. In this arrangement, each cell will produce current at a voltage limited by its bandgap, which will not normally be much lower than photon energy absorbed by each cell as illustrated schematically in Fig. 16. The reflectors noted in Fig. 16 prevent luminescent radiation produced in each solar cell from passing into the succeeding cell where it would be converted with lower efficiency. Luque (2011) has indicated that even for an infinite number of ideal cells as in Fig. 16, the efficiency would be 86.8 % and below the Landsberg efficiency expressed in Eq. 1. Stacks of solar cells as in Fig. 16 are referred to as *multijunction solar cells*, and a number of so-called monolithic multijunction solar cell stacks have been fabricated. These usually consist of solar cells of different materials along with

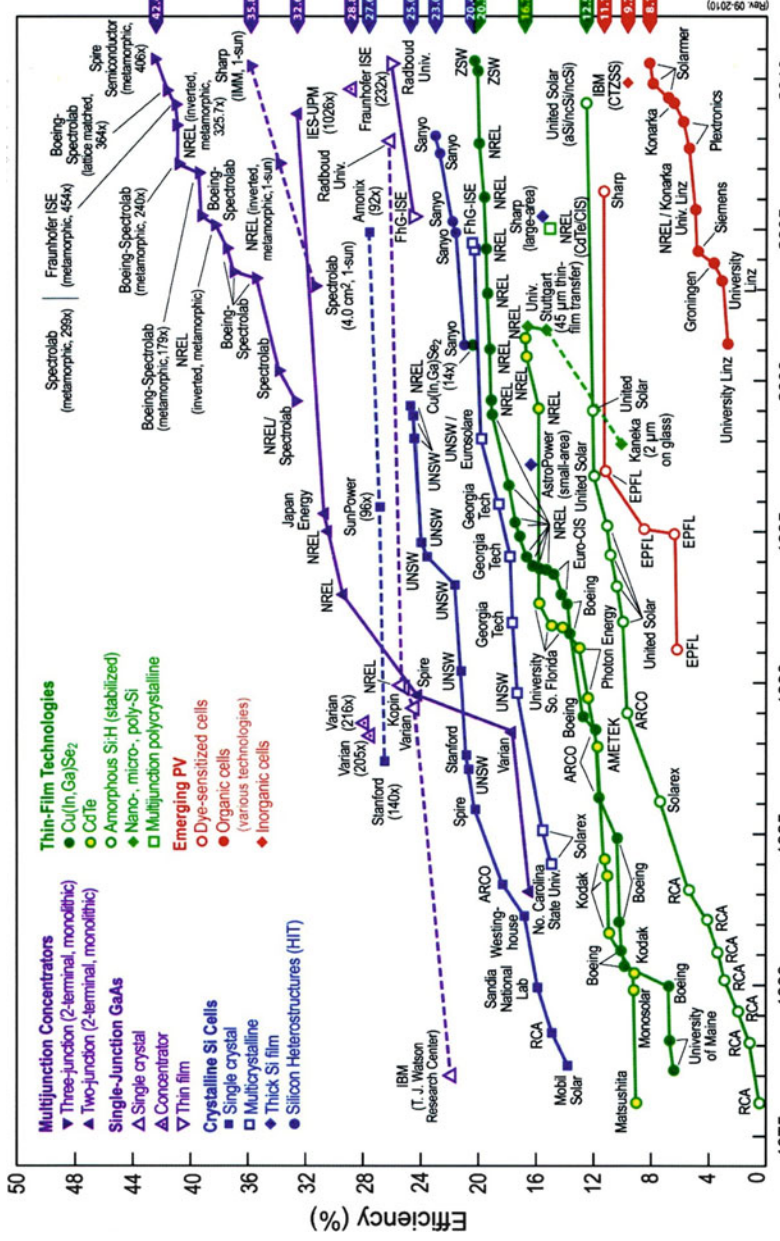


Fig. 14 Trending efficiencies for experimental (research) and some commercial solar (photovoltaic) cells from 1975 to 2010 (Courtesy of the U.S. National Renewable Energy Laboratory (NREL), Golden, CO)

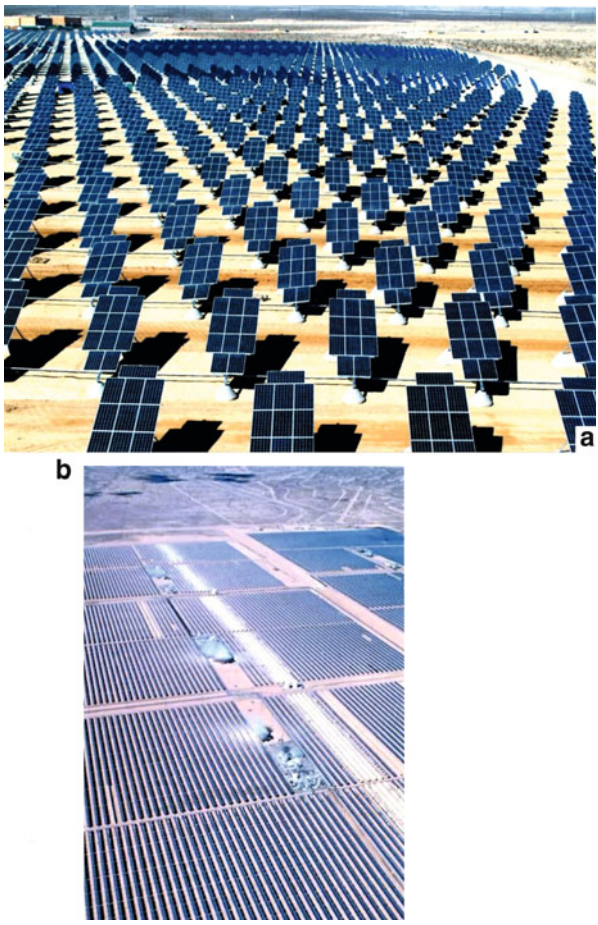


Fig. 15 Examples of solar power plants consisting of solar-photovoltaic conversion panel arrays. (**a**) *p*-type and *n*-type Si photovoltaic solar cell arrays generate 15 MW of solar electrical power for Nellis Air Force Base, Nevada, USA, on 57 ha of unused land. (US Air Force photograph). (**b**) Kramer Junction (Mojave Desert, CA) solar electric generating station: 150 MW output (Courtesy of FPL Energy)

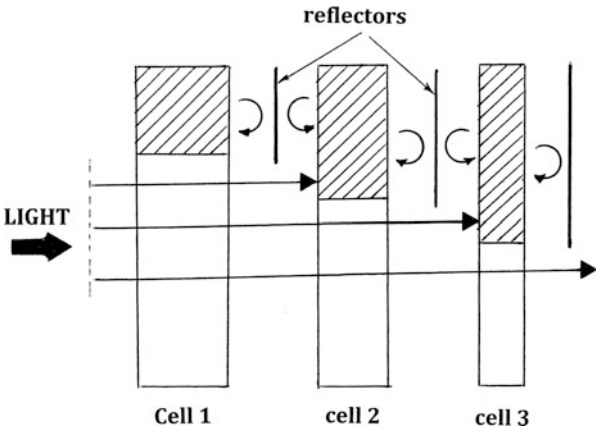


Fig. 16 Stack of solar cells with decreasing bandgap or bandgap energies – from *left* to *right* (Adapted from Marti and Araujo (1996))

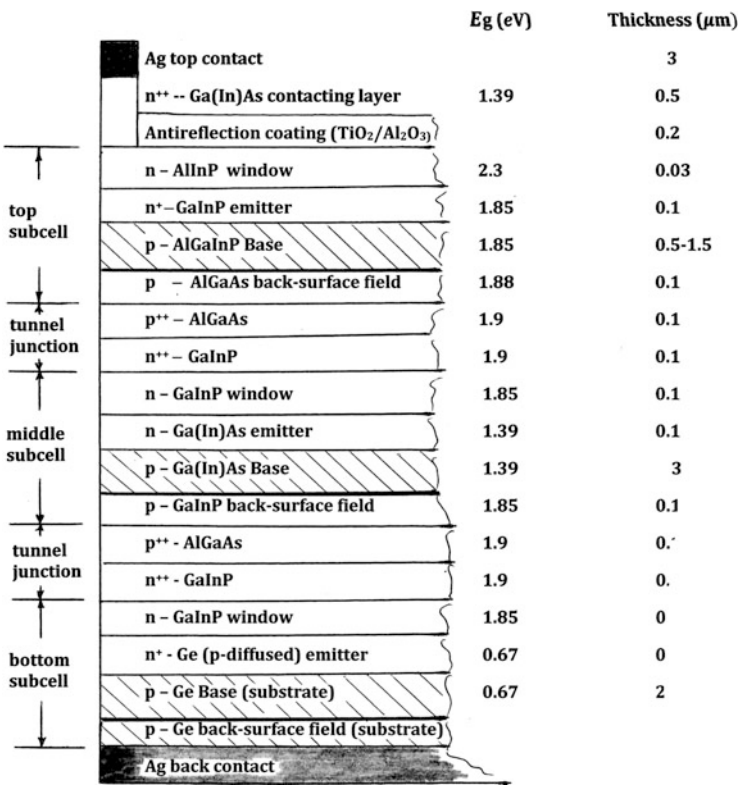


Fig. 17 Schematic cross section of a series connected, three-junction solid-state solar cell (Fig. 16). Energy (band)gap and layer thickness are indicated. Doping indicated by n^{++}, n^{+}, and n or correspondingly p^{++}, p^{+}, and p corresponds to electron or hole concentrations on the order of 10^{19}–10^{20}, 10^{18}, and 10^{17} cm^{-3}, respectively (Adapted from Friedman et al. (2011))

tunnel diodes to connect them in series, and all grown on the same or single substrate. The semiconductor junctions are epitaxially grown with a similar lattice constant so that there is no mismatch, or little mismatch.

Solar cell stacks represented ideally by three cells in Fig. 16 have been fabricated on germanium substrates as illustrated schematically in Fig. 17. Germanium matches the lattice parameter of III–V compound semiconductors grown epitaxially in fabricating such cell stacks. While efficiency limits of ~63 % have been determined for these configurations (Fig. 16), efficiencies of ~42 % have been realized in limited commercial products. Although organic multijunction arrangements are considerably less efficient, their production costs are considerably less than solid-state arrays and organic multijunction devices are applicable in large-area, flexible system production, especially roll-to-roll 3D printing concepts. The stacked arrays as shown schematically in Fig. 17 are expensive and rigid, but relatively efficient.

A low-cost, novel, thin-film, flexible solar cell which has achieved conversion efficiencies of around 15 % involves the dye or *dye-sensitized solar cell*

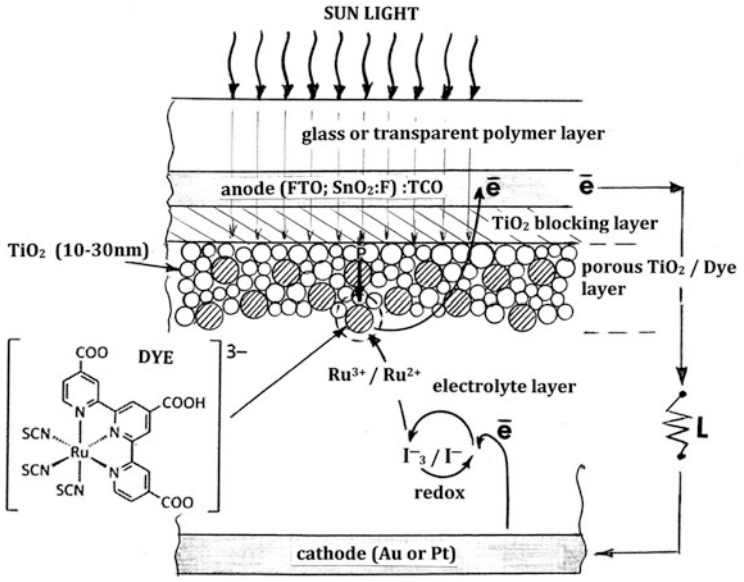


Fig. 18 Schematic showing dye-sensitized solar cell structure and device operation. The dye illustrated is an anionic Ru-terpyridine complex (shown) which is referred to as Black Dye. This molecule absorbs a photon (p) and the sunlight passes through the transparent electrode (*dotted circle*) elevating the Ru^{3+} to $Ru^{2+} + \overline{e}$ and the electron $(\overline{e})$ injected into the TiO_2 conduction band percolates through the porous TiO_2 layer to the anode, where it passes through the circuit load (L) to the cathode and transferred through the electrolyte to triodide to yield iodide (the redox cycle). The iodide reduces the oxidized Ru^{3+} to Ru^{3+} (the original valence state). TCO represents the transparent conducting electrode, usually fluorinated SnO_2. The electrolyte is a liquid conductor containing an iodide $\left(I_3^- + 2\overline{e} \rightarrow 3I^-\right)$

(DSC or DSSC) discovered half century ago when illuminated organic dyes were observed to generate a voltage at oxide electrodes in electrochemical cells. However, the contemporary dye-sensitized solar cell, sometimes called the Grätzel cell (or Grätzel dye cell), coinvented by Brian O'Regan and Michael Grätzel (O'Regan and Grätzel 1991), is based on photosynthesis biomimetics originally described by H. Tributsch and M. Calvin (Tributsch and Calvin 1971) who received the 1961 Nobel Prize in Chemistry, "for research on the carbon dioxide assimilation in plants," or the fundamentals of photosynthesis. Grätzel was able to simulate the primary processes involved in photosynthesis by fabricating a porous layer of TiO_2 nanoparticles covered with a molecular dye that absorbs sunlight in the same way chlorophyll absorbs sunlight in plant leaves. The configuration emulates a conventional alkaline battery consisting of the transparent TiO_2 anode and a conducting metal cathode (usually Au or Pt) which sandwich an electrolyte containing the organic dye. Sunlight passes through the semitransparent TiO_2 layer and excites electrons (or holes) in the dye layer electrolyte which then flow into the TiO_2. Figure 18 illustrates the dye-sensitized solar cell structure. Many variations of this structure continue to evolve, including the electrodes and electrolytes (or organic dye

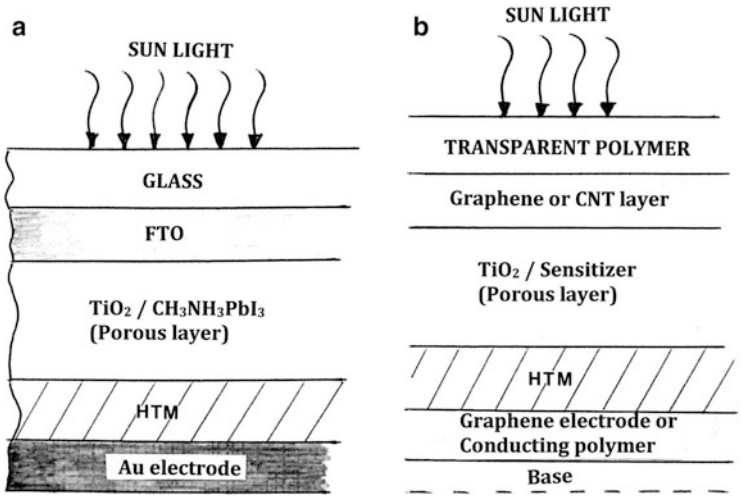


Fig. 19 Schematic diagrams for innovative dye solar cell modules. (**a**) DSSC using hybrid perovskite $CH_3NH_3PbI_3$ dye formed in nanoporous TiO_2 film: perovskite-based photovoltaic device structure. (**b**) Flexible, thin-film DSSC structure using graphene electrodes. The HTM layer in (**a**) and (**b**) is the hole-transporting material: *p*-type polymer. The mesoporous TiO_2 films are colloidal anatase (nanoparticles) deposited onto a compact TiO_2 underlayer deposited by spray pyrolysis onto the FPO film (**a**) or graphene (**b**)

and dye sensitizers) which allow for greater solar spectral sensitivity such as the anionic Ru-terpyridine complex shown in Fig. 18 (referred to as Black Dye) (Grätzel 2003), copper–diselenium (Cu (In, Ga) Se_2) which has a higher conversion efficiency, and the more recent hybrid perovskite-sensitized mesoscopic solar cells produced by sequential deposition (Burschka et al. 2013). Figure 19 illustrates this configuration and other innovations involving graphene and carbon nanotube (CNT) electrodes which are being explored. These thin-film configurations can be fabricated in flexible arrays using 3D printing (roll-to-roll) processing and can attain conversion efficiencies exceeding 15 %. In addition, it is certainly possible to fabricate stacked arrays or DSCs as illustrated in Fig. 16 in order to significantly improve their overall conversion efficiency and cost-effectiveness, even though such stacks may involve single-bandwidth cells as illustrated in Fig. 19.

The implications in Fig. 19 are also that such cells may include metal oxide–polymer hybrid solar cells where the TiO_2 is used as a thin *n*-type electrode, or alternately ZnO. The metal oxide can alternately be coated with an inorganic absorber like crystalline CdSe or Sb_2S_3.

Concentrator technologies are also used increasingly to assure a high luminous flux which is necessary to achieve ideal solar cell efficiencies. The simplest form for such concentrators involves a lens, especially a Fresnel lens with the shape of a dome. These can be placed on small solar cell modules or a parquet of several Fresnel lenses covering larger modules (Sala and Anton 2011). Other novel concepts for optimizing

solar cell efficiencies involve growing nanoscale structures or nanowire or rodlike features on the solar cell module surface to capture more of the incoming sunlight. Conversion of heat or waste heat recovery during photovoltaic conversion is also a viable augmentation to solar energy conversion as implicit in fabricating hybrid photovoltaic–thermoelectric devices which will be discussed later in this chapter.

Thermoelectric Materials and Device Applications and Innovations

In 1821, a German physicist named Thomas Seebeck discovered that a compass needle was deflected by a closed loop formed by two different metals joined in two places, with a temperature difference between the metal junctions. Known as the *Seebeck effect*, this is the principle for thermocouples. A little more than 10 years later, Jean Charles Peltier discovered the presence of heating or cooling at an electrified junction of two different conductors which became known as the *Peltier effect*. For each material, the cooling effect is given by the *Peltier coefficient*, which relates the heat carried by the changes to the electrical current (I):

$$Q = \Pi \times I \tag{3}$$

The corresponding power generation is measured by the *Seebeck coefficient*, S, which relates the voltage generated to the temperature difference through

$$\Delta V = S \Delta T \tag{4}$$

The Peltier and Seebeck coefficients are related through the Kelvin relation:

$$\Pi = ST \tag{5}$$

Typically, metals have small Seebeck coefficients because they have partially filled energy bands, and their conductivity, σ, is insensitive to small changes in energy. Energy and entropy transport accompany the motion of charges in conductors. In contrast, semiconductors can be doped (n or p type) to manipulate excess electrons or holes which can allow the value of the Seebeck coefficient, S, to be varied over a range of positive (+) or negative (−) values; this sign can be used to determine whether electrons (−) or holes (+) dominate the charge transport in a semiconductor.

As shown in the p-n junction systems in Fig. 20, simple Seebeck circuits can be configured as either a thermoelectric generator (Fig. 20a) or a thermoelectric (Peltier) cooler, or refrigerator (Fig. 20b). These simple circuit arrangements constitute a thermoelectric device with *n*-type and *p*-type legs electrically connected in series and thermally in parallel. The electrons in the *n*-type material and holes in the *p*-type material carry heat away from the cooled metal–semiconductor junction in Fig. 20b, while conversely if a temperature difference is maintained between the two ends of the semiconductor materials as in Fig. 20a, higher thermal energy electrons and holes diffuse to the cool side and create a potential difference (voltage) across the external load (resistance).

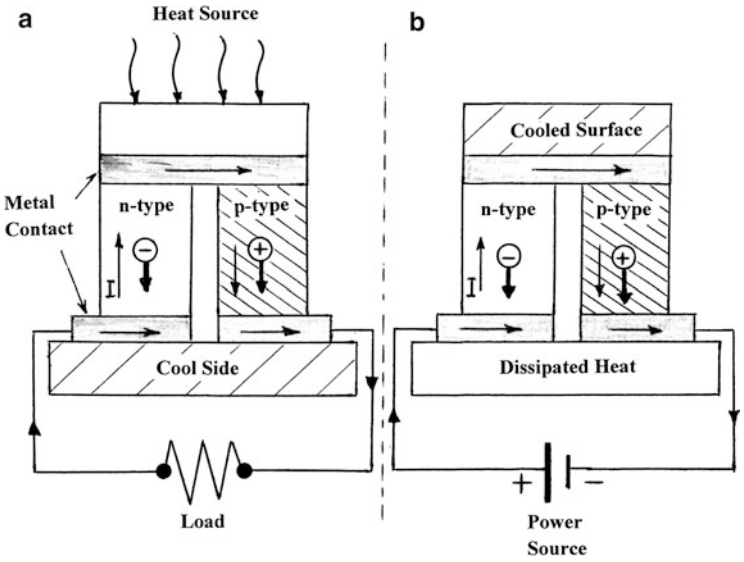


Fig. 20 Simple Seebeck circuit as a thermoelectric generator (**a**) or a thermoelectric cooler (**b**). The Seebeck coefficient, S, is different for the *n*-type material in contrast to the *p*-type material. Current flow is denoted I

Practical thermoelectric devices, thermoelectric generators or thermoelectric refrigerators, are made from multiple pairs of *n*-type and *p*-type semiconductor legs because putting these elements electrically in series and thermally in parallel increases the operating voltage of the module while reducing its current. Figure 21 illustrates this configuration for a thermoelectric generator (TEG) module. The heat balance equation as described by Goldsmid (1964) shows that efficient thermoelectric power generators and refrigerators should have a large *figure of merit* defined by

$$ZT = \sigma S^2 T/K \tag{6}$$

where σ is the electrical conductivity, S is the Seebeck coefficient (or thermoelectric power), T is the temperature, and K constitutes the electronic and lattice contributions to thermal conductivity:

$$K = K_e + K_L \tag{7}$$

The ability of a material to produce useful electrical power under a given thermal gradient is given by its *power factor* $= \sigma S^2$. It can be observed from Eq. 6 that to achieve a high ZT requires a high Seebeck coefficient, S, and electrical conductivity, σ, and a correspondingly low thermal conductivity, K. Additionally, for an actual thermoelectric module as shown in Figs. 20 and 21, two materials (n and p type) are used which results in a modified, dimensionless figure of merit:

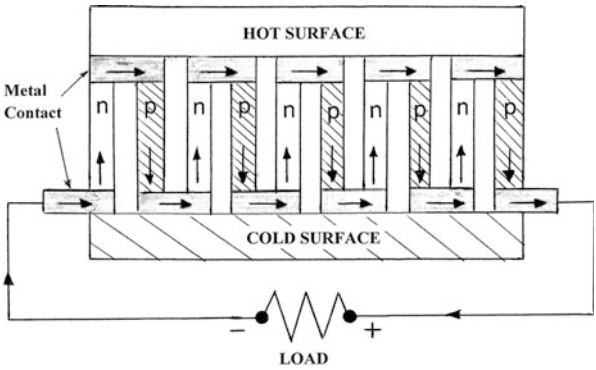


Fig. 21 *n-p*-type series module schematic for thermoelectric generator (TEG). *Arrows* show current flow opposite to electron flow in *n*-type and in same direction as hole flow in *p*-type material

$$Z\overline{T} = \left(S_p - S_n\right)^2 \overline{T} / \left[\left(\rho_n K_n\right)^{1/2} + \left(p_p K_P\right)^{1/2} \right]^2 \tag{8}$$

where S_p and S_n are the corresponding *p*-type and *n*-type Seebeck coefficients, ρ_p and ρ_n are the electrical resistivities ($\rho = 1/\sigma$), and $\overline{T}$ is the average temperature between the hot and cold surfaces; K_p and K_n are given for each material as expressed in Eq. 7.

Over the past half century, bismuth chalcogenides such as Bi_2Te_3 and Bi_2Se_3 and alloys such as $(Bi_{0.5}\,Sb_{1.5})\,Te_3$ have been the basis for TEG modules and solid-state refrigeration systems with ZT $\sim$ 1 at 300 K (Goldsmid 1964). Like superconductors prior to the discovery of high-temperature oxide superconductors in 1986 (chapter "▶ Structures and Properties of Oxide Superconductors"), even small decimal point increases in the dimensionless figure of merit (ZT) have been significant. And like superconductors, there is no adequate theory for thermoelectric phenomena which can provide clear strategies for producing high-efficiency thermoelectric materials and devices. However, in the early 1990s, Hicks and Dresselhaus (1993a, b) proposed that ZT could be increased by preparing thermoelectric materials in 2D or 1D nanogeometries or quantum confinements. In addition to electron confinement, such nanoconfigurations should reduce the phonon thermal conductivity K, or K_L. They performed calculations on 3D- (bulk) 2D-layered or quantum well superlattice structures (Hicks and Dresselhaus 1993a) and a 1D quantum wire (Hicks and Dresselhaus 1993b) of Bi_2Te_3 and produced plots of ZT versus layer thickness and quantum-wire diameter, a, reproduced in Fig. 22. As observed in Fig. 22a, Bi_2Te_3 quantum well structures had the potential to increase ZT by a factor of 13 over the bulk value, while as observed in Fig. 22b, ZT of up to a factor of 14 was predicted for a wire of square cross section and width of 5 Å. These predictive results indicated that quantum well and wire structures had the potential to significantly improve the thermodynamic figure of merit.

While the original predictions by Hicks and Dresselhaus (1993a, b; Fig. 22) do not rigorously apply to quantum dots, the enhanced power factor (σS^2) in 1D and 2D quantum confinement structures is optimized in the direction perpendicular to

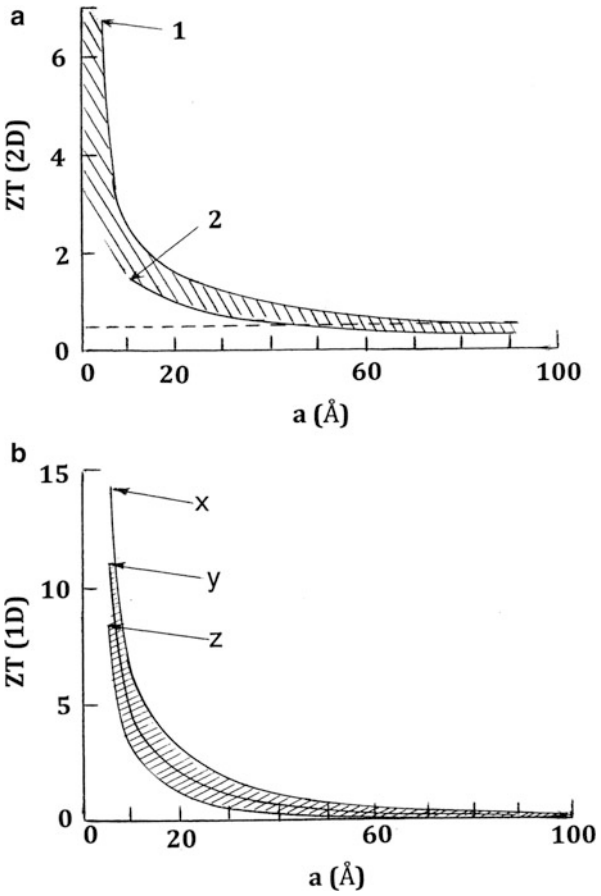


Fig. 22 Plots of SZT versus layer thickness, a, for a–b (*1*) and a–c (*2*) plane layers (**a**) and the width (**a**) of 1D wires fabricated along the x, y, and z directions (**b**). The *dashed line* in (**a**) indicates the best ZT for 3D (bulk) Bi_2 Te_3 thermoelectric material. a, b, and c represent the crystallographic directions where **c** is perpendicular to the a–b planes and **b** is perpendicular to the a–c planes. Shading represents the range of values (Adapted from Hicks and Dresselhaus (1993a, b))

the confinement (Shakouri 2011). Cai and Mahan (2008) recently showed that quantum dot arrays pose potentially higher ZT for thermoelectric materials. Li et al. (2010) have also recently demonstrated that nanostructures of various forms, nano-dispersions, and nanoscale heterogeneities are effective in reducing the thermal conductivity to a greater degree than electrical conductivity, resulting in increased ZT for bulk, thermoelectric materials. There is a significant decrease in K_L (Eq. 7) due to scattering at nanograin boundaries and nano-precipitates, especially where the grain or phase size is smaller than the phonon mean free path. Such scattering phenomena are also associated with dislocations and other crystal defects, especially interfaces (Pichanusakorn and Bandaru 2010; Rowe 2006).

Variations in Bi_2Te_3 and alloy variant fabrication have pushed ZT to values above 2 (at room temperature), but efficient device structures have been manufactured for decades with ZT $<$ 1. PbTe and Th-doped PbTe alloy have achieved ZT $>$ 1 at $\sim$775 K. $SrTiO_3$-based materials and $CaMnO_3$-based materials (especially layered perovskite-type structures) have exhibited promising thermoelectric behavior with a

high power factor (σS^2), but the lattice thermal conductivity (K_L) is generally too high (Koumoto et al. 2010). Oxide thermoelectric modules such as *p*-type $Ca_{2.7}$ $Bi_{0.3}$ Co_4 O_9 and *n*-type $Ca_{0.9}$ $Yb_{0.1}$ MnO_3 using Ag electrodes have been fabricated (Koumoto et al. 2010). Dragoman and Dragoman (2007) have also shown that a very high Seebeck coefficient (~30 mV/K) at room temperature (in contrast to ~1 mV/K for Se and 0.5 mV/K for Te) might be achieved in graphene, with a corresponding ZT of ~20. Similarly, Rowe (2006) has described theoretical studies where organic conductors may have figures of merit approaching 20, especially in the form of so-called quasi one-dimensional organic crystals formed from linear chains or stacks of molecules which compose 3D polymer crystals.

Thermoelectric power generation is one of the promising methods to recover various forms of natural (sunlight) heat as well as many sources of waste heat, especially if thermoelectric materials with $ZT > 3$ can be produced and configured into low-cost modules with reasonable conversion efficiencies approaching those of common photovoltaics (~10–15 %). The ability to produce large-area thermoelectric generator module configurations, especially printable, and flexible device configurations could have an enormous impact on electricity generation. Novel concepts could include electricity generation from warmed (hot) asphalt highways as illustrated conceptually in Fig. 23. Similar flexible module arrays could also be wrapped around hot pipes and other waste heat configurations.

Another innovation for solar energy conversion is the fabrication of *photovoltaic/thermoelectric hybrid modules* illustrated schematically in Fig. 24. This would allow electricity generation from a larger portion of the solar energy spectrum along with the waste heat generated from solar-facing photovoltaic arrays shown in Fig. 15. Park et al. (2013) fabricated a module shown schematically in Fig. 24 using *n*- and *p*-type (doped) crystalline Si in the PV device and commercial Bi_2Te_3 and SbTe *n*- and *p*-type legs in the TE array. The conversion efficiency and power output of the PV device was measured to be ~30 % at a 15 °C temperature gradient across the TE device. Figure 25 illustrates a commercial TEG module construction. Modules of ~6 cm^2 and 2–5 mm thick as in Fig. 25 can generate over 100 mV/°C.

It might be noted that most semiconductor materials, and especially those used in PV devices, loose efficiency at the rate of ~5 % for every 10 K (Kelvin) increase in temperature, and this can compromise hybrid module use in very hot regimes. Designs employing wavelength separators (or selective sunlight absorbers) that transmit the UV and visible wavelengths to the PV cells while guiding the IR-radiation to the TE generator module can alleviate this efficiency degradation.

Exhaust gas waste heat energy recovery in hybrid electric vehicles (HEVs) is a viable application not only for TEG innovations but also photovoltaic–thermoelectric hybrid energy sources. These concepts have recently been discussed by Zhang et al. (2009).

It might be observed in retrospect that optoelectronic devices generally benefit from structuring in multiple length scales, although preserving crystal epitaxy as a requisite for good optoelectronic properties, while creating a complex, three-dimensional structure continues to pose a significant materials science and engineering challenge. Manipulating microstructures as these relate to length scales to optimize device performance is complicated in hybrid module fabrication since

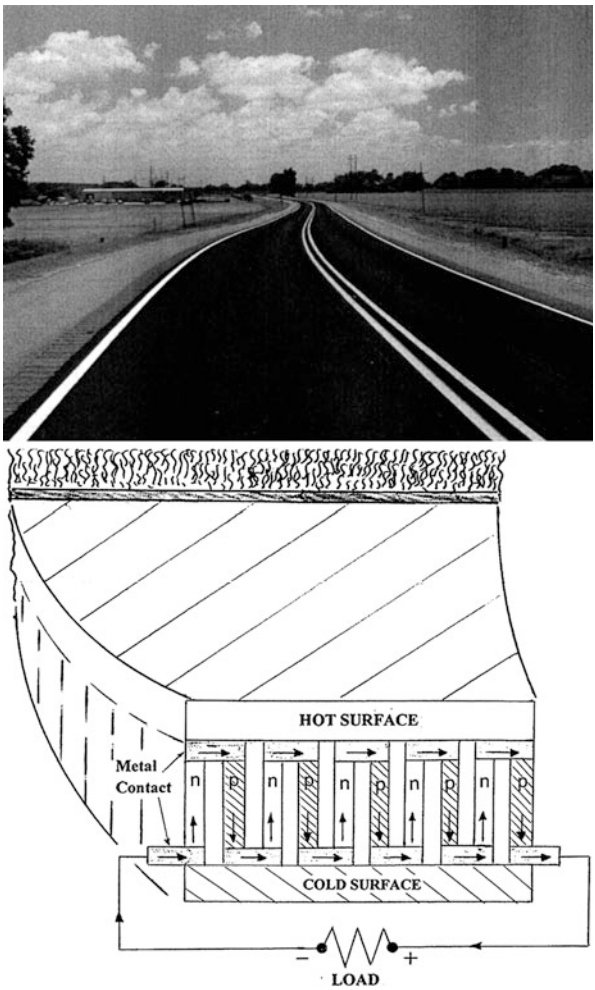


Fig. 23 Integration of large-area thermoelectric module arrays (Fig. 21) into asphalt highway construction for electricity generation

length scales which can optimize thermoelectric behavior, for example, compromise electronic conduction, and as noted earlier, the interactions or interdependencies of the many issues relating to device efficiency are largely unknown. In the long run, it may suffice to fabricate cheap, sustainable, large-area arrays with relatively low conversion efficiencies. The ability to print such arrays on flexible substrates by 3D printing may provide important technology breakthroughs.

Battery Technologies and Energy Storage Materials Innovations

Because of a global interest in electric vehicles (EVs), battery development has become a major thrust of materials development and innovation. Electric vehicle production actually began long before the proliferation of internal combustion

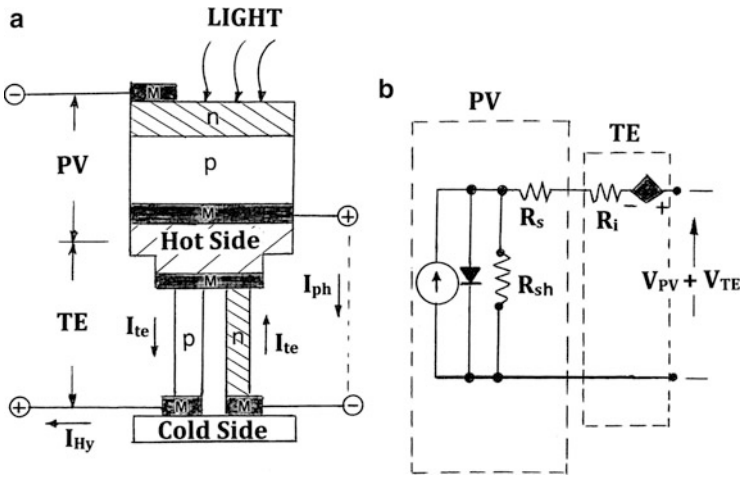


Fig. 24 Photovoltaic (PV)–thermoelectric (TE) hybrid module. (**a**) Schematic view of PV-TE device electronically connected in series. Photo-generated current and Seebeck (TE) current are denoted I_{ph} and I_{te}. Output current in hybrid circuit is denoted I_{Hy}. Metal contact conductors are denoted M. (**b**) Equivalent electric circuit diagram. The TE module is represented by a voltage source connected with an internal (series) resistance, R_i. The PV cell is a current source shunted by a diode with a series, resistance, R_s, and a shunt resistance, R_{sh} (Adapted from Park et al. (2013))

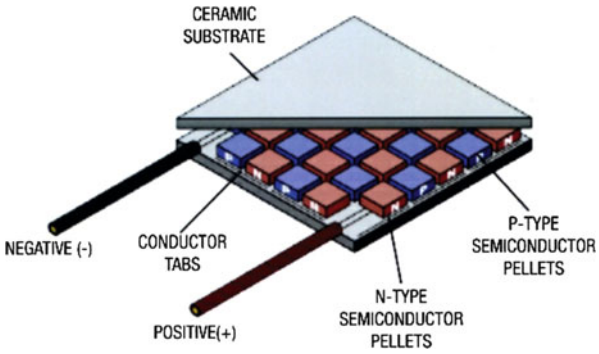


Fig. 25 Thermoelectric generator (TEG) module structure (Courtesy of Lineai Technology)

vehicles: around the middle of the 1800s in the USA, and by 1900, nearly one-third of all automobiles on US highways were electric, with a range of up to 50 miles per battery charge. But by 1913, mass production of cheaper and longer-range gasoline-powered automobiles put EVs in decline. In 2001, Toyota introduced the first mass-produced hybrid gasoline-electric car. EV production has been rising in the past decade, but cost, range, and changing infrastructure remain a detriment to public acceptance of automotive EVs or battery electric vehicles (BEVs). However, the Tesla Roadster introduced in 2008 in the USA was the first production vehicle to use lithium-ion battery cells which allowed for travel ranges of slightly more than 240 miles (320 km) on a charge. This vehicle can accelerate from 0 to 60 mph in

less than 4 s and achieve a top speed of ~120 mph. Recharge times for a range of ~200 miles was ~4 h for the lithium-ion battery pack. Consequently, the key to public acceptance of BEVs will be a dramatic reduction in vehicle cost (and battery cost), doubling or more in the travel range (up to 500 miles in the USA), and readily available charging stations. Consequently, the materials challenge is primarily in the dramatic improvement in lithium-ion battery performance utilizing materials or material component innovations, or the development of novel approaches to vast improvements in battery technology. Lithium–air batteries which use the oxidation of Li at the battery anode and reduction of oxygen at the cathode to induce current flow have the potential of up to 15 times the current Li-ion battery energy density, which would more than triple the current Li-ion battery capability and efficiency. For comparison, the energy density (or specific energy) of gasoline is ~13 kWh/Kg in contrast to a theoretical energy density of ~0.8 kWh/Kg for Li-ion batteries. Consequently large battery packs containing more than 100 battery units are required in automotive EVs.

In contrast to EVs and BEVs, there is a huge demand for much smaller battery operation and efficiency in the consumer, portable electronic sector worldwide. Here miniaturization and reduced costs drive innovations in solid-state battery fabrication, including thin-film and flexible battery platforms as well as light weighting of batteries and battery packs and arrays.

The basic unit comprising any battery is an electrochemical cell which involves conducting charged particles – electron and ions – from a cathode to an anode in a primary cell or from anode to cathode in a secondary cell. As illustrated generally in Fig. 26a, electrons are transferred from the negative electrode (anode) to the positive electrode (cathode) during discharge through a load resistor. Correspondingly, during charging, electrons flow in the opposite direction, and the position of the anode and cathode is reversed. During discharge, positive ions (cations) flow from the negative electrode to the positive electrode, in the same direction as the electrons. However, during charging (or recharging), the electrons and ions flow in the opposite direction. Correspondingly, current flows opposite to the flow of electrons (and ions).

Table 1 compares a few contemporary battery systems and component materials for liquid-state electrolyte batteries. The electrolyte serves as the conduction medium for ion transport from one electrode to the other (an ionic conductor). The Ni–Cd- and Ni–MH-type rechargeable batteries use positive electrodes of nickel hydroxide or nickel oxyhydroxide, but Ni–MH cells use a hydrogen-absorbing alloy instead of Cd, although Ni–Cd battery technology is largely obsolete. The electrolyte in both cells is KOH. The corresponding electrode reactions for the Ni–MH cell are illustrated in the footnote of Table 1 for discharge (reaction arrow to left) or charge (reaction arrow to right).

For the Li-ion cell, and considering Fig. 26a, the positive electrode half reaction is

$$LiCoO_2 \leftrightarrows Li_{l-n}CoO_2 + n\,Li^+ + n\,\overline{e} \tag{9}$$

while the corresponding negative electrode half reaction is

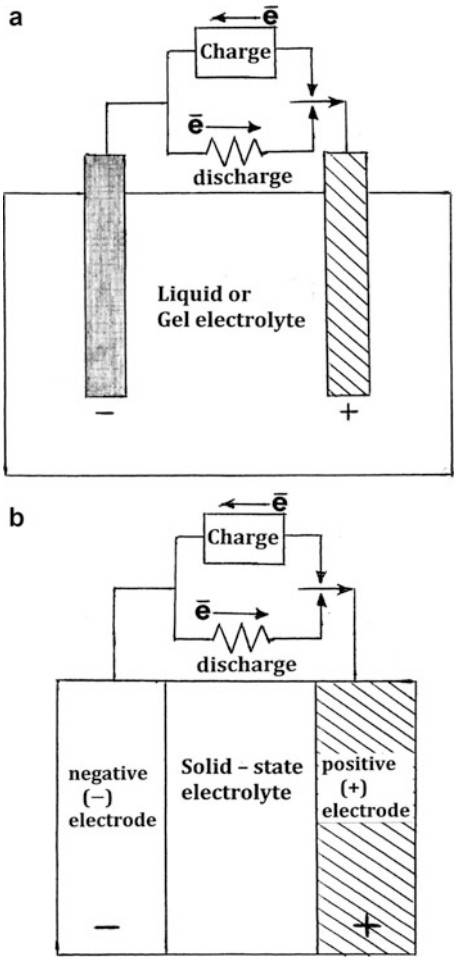


Fig. 26 Electrochemical (battery) cell discharge–charge cycles. (**a**) Liquid or gel electrolyte. (**b**) Solid-state electrolyte

$$nLi^{+} + n\bar{e} + C \leftrightarrows Li_nC \tag{10}$$

where the reaction arrow directions are consistent with those shown in Table 1 footnote for the Ni–MH cell. Optimization of electrical (electron) and ionic conductivities in cathode (positive electrode) materials is important since either of these can influence overall cell properties, including the cell capacity and charge–discharge cycle life. Consequently, cathode microstructure as this relates to diffusivity is important, and compared to Li-ion conduction, electrical conduction is not a dominant factor.

Typical Li-ion cells use carbon (graphite) as the anode, and conduction properties are complex and dependent upon the degree of crystallinity of the graphite as well as the porosity or related electrode surface area. Both the cathode and anode porosities and surface areas are important issues in the overall cell operation.

Table 1 Comparison of contemporary battery types

Battery system	Negative electrode (−)	Positive electrode (+)	Electrolyte	Nominal voltage (V)	Theoretical specific energy (Wh/kg)	Practical specific energy (Wh/kg)	Practical energy density (Wh/L)
Lead–acid	Pb	PbO_2	H_2SO_4	2.0	252	35	70
Ni–Cd	Cd	NiOH or NiOOH	KOH	1.2	244	50	75
Nickel MH (MH: metal hybrids)	MH[a]	NiOH or NiOOH	KOH	1.2	278–800	70–120	170–300
Li-ion	Li	Li_xCoO_2	PC or DMC w/Li PF_6[b]	4.0	766	120	200
Li–air	Li	O_2	Carbonate-base organic[c]	2.9	11,140	–	–
Na–S	Na	S	B–Al_2O_3	2.0	792	100	>150

[a]M is an intermetallic, commonly AB_5 where A is a rare earth and B is Ni, Co, Mn, and/or Al. These reversibly form a mixture of metal hydride compounds: $H_2O + M + \overline{e} \leftrightarrows OH^- + MH$. At the positive electrode the reaction is: $Ni(OH)_2 + OH^- \leftrightarrows NiO(OH) + \overline{e}$

[b]Polycarbonate (PC), dimethyl carbonate (DMC) or mixture. A typical electrolyte is Li PF_6 dissolved in a mixture of PC and DMC

[c]Aprotic (non-proton donor) and aqueous electrolytes can produce varying electrochemical reactions at the cathode

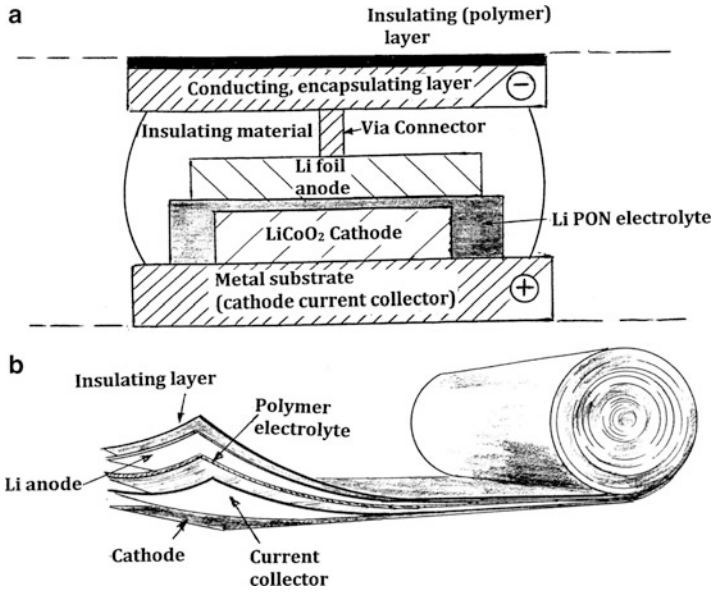


Fig. 27 Solid-state, thin-film-layered Li-ion battery concepts. (**a**) LiPON electrolyte cell schematic. (**b**) Li–polymer battery

Liquid electrolytes in Li-ion batteries consist primarily of $LiPF_6$ in organic solvents such as ethylene carbonate, dimethyl carbonate, or diethyl carbonate. Other Li salts such as $LiClO_4$ and $LiBF_4$ are also used. The electrolyte conductivity is also an important issue. Liquid electrolytes pose problems in extreme temperature regimes as well as containment or leakage issues.

For many battery application configurations, solid electrolytes have a significant advantage. This is especially true for thin-film battery fabrication as well as large storage arrays in transportation (automotive) applications. A common solid-state electrolyte in Li-ion cells (Fig. 26b) involves the use of solid polymers in place of solvent matrices. Other solid-state electrolytes include lithium phosphorous oxynitride (Li_2PO_2N) or LiPON($Li_xPO_yN_z$) which allows for high Li^+ conduction (Senevirathne et al. 2013). This material can be conveniently sandwiched between a $LiCoO_2$ cathode and a Li metal (foil) anode as illustrated schematically in Fig. 27a, especially in constructing micro-batteries. Alternatively, Li foil anode configurations with flexible polymer-matrix electrolyte layers and thin foil cathode materials such as graphene could allow large-area battery structures to be fabricated as illustrated schematically in Fig. 27b. LiI/Al_2O_3 electrolyte mixtures can also be effective. The use of lithium metal as the anode in rechargeable cells as opposed to carbon materials in conventional Li-ion batteries can dramatically increase energy density in solid-state Li-ion battery configurations shown in Fig. 27. In addition, micro-battery arrays can be fabricated by 3D printing technologies utilizing the layer concept illustrated in Fig. 27a and printed onto large-area formats suggested in Fig. 27b.

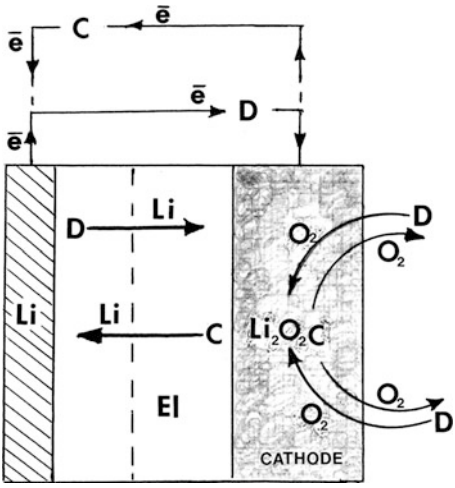


Fig. 28 Li–air battery charge. (C) discharge (D) cycle schematic design. Electrolyte, aqueous, aprotic, aqueous/aprotic, and solid state, is denoted El. Electron flow is denoted $\bar{e}$. The *dashed line* represents a conducting membrane for aqueous/aprotic and solid-state electrolytes

One of the more recent battery technology innovations involves the lithium–air battery that uses lithium oxidation (by air) at the anode and oxygen reduction at the cathode to create an electrical current. As illustrated in Table 1, Li–air battery energy density is comparable to traditional gasoline (as noted earlier). Other metal–air (oxygen) batteries such as Al and Zn anode batteries have also been developed over the past decade (Li et al. 2013). Lithium–air batteries theoretically have the potential to store up to 50 times more energy than more conventional Li-ion batteries, a particular advantage in automotive electric vehicle (BEV) applications.

Figure 28 illustrates a general schematic for Li–air battery charge–discharge cycles which is essentially the same electrochemical cell configuration as Fig. 26, with Li^+ ions flowing to the right in discharge (D) mode and to the left in charge (C) mode. Correspondingly, air (or oxygen, O_2) flows into the cathode (positive electrode) in the discharge (D) mode and out of the positive electrode in the charge (C) mode. Mesoporous carbon has been used as the cathode material with metal catalysts which enhance oxygen reduction kinetics (Besenhard 1999). These include Co, Pt, Ag, and Mn which seems to be favored. Electrolytes take four forms: aqueous, aprotic, aqueous–aprotic, and solid state. For aqueous electrolytes, which combine lithium salts dissolved in water, there is no Li^+-conducting membrane shown in Fig. 28. The aprotic type Li–air battery design employs an aprotic electrolyte composed of an organic solvent with a dissolved lithium salt rather than water. Typically carbonates, esters, etc., with dissolved $LiPF_6$ and $LiSO_3$ CF_3 are used. The cathode is usually highly porous carbon impregnated with a nanometal oxide such as MnO_2 or Mn_3O_4.

The mixed aqueous/aprotic electrolyte Li–air battery consists of a 2-part electrolyte, one aqueous and one aprotic, separated by a lithium-conducting membrane (dashed partition in Fig. 28). The aprotic electrolyte is in contact with the Li anode, while the aqueous side of the electrolyte is in contact with the porous cathode. The Li^+-conducting membrane separation is often a thin Li^+-conducting ceramic membrane.

The solid-state Li–air battery design is similar to the hybrid aprotic/aqueous cell with a Li^+-conducting ceramic membrane separating a polymer-ceramic composite electrolyte which enhances charge transfer at the anode while electrochemically coupling the cathode to the solid electrolyte. The solid-state Li–air battery like other solid-state battery designs enables monolithic fabrication of thin-film structures illustrated conceptually in Fig. 27 (Kumar et al. 2010).

As noted previously for Li-ion battery designs, the cathode structure or microstructure can be manipulated to increase the surface area for reaction and the incorporation of catalytic materials, especially nanocatalysts which also have large surface area for reaction. Catalysts can assist in Li_2O_2 nucleation which is irreversible with respect to Li. Electrolyte development is also a challenge in solid-state Li–air battery designs since glass-ceramic electrolyte compositions pose Li^+ conductivity issues. Cathodes composed of graphene intercalating catalyst nanoparticles and related nano-layer compositions may prove to be promising innovations in Li–air solid-state battery designs in particular.

Supercapacitors

Simple capacitors (electrostatic capacitors) discussed generally in Part II, chapter "Electromagnetic Fundamentals" (Fig. 9), conduct alternating current (AC) and block direct current (DC). They can also store charge, or energy embedded in the stored charge. Unlike simple dielectric (ε_r) capacitors, electrolytic capacitors are polarized and have much larger capacitance than dielectric capacitors. Electrolytic capacitors usually have two aluminum plates where an Al_2O_3 layer acts as the dielectric (with relative dielectric constant, $\varepsilon_r \sim 10$). An electrolyte is inserted between the electrode plates, usually consisting of an aqueous solution of boric acid or sodium borate, along with substances such as ethylene glycol to decrease evaporation. Tantalum electrodes are also used in electrolytic capacitors, but Ta is more expensive and not as easily oxidized, although they (oxides) are more stable. Solid (conductive polymer) electrolytes have also been developed for both aluminum and tantalum electrolytic capacitors.

In contrast to electrostatic (dielectric) capacitors and electrolytic capacitors, supercapacitors (also called ultracapacitors) or electric double-layer capacitors (EDLCs) characterize a new type of electrochemical capacitor where the capacitance is characterized by double-layer capacitance, electrostatic storage which results by charge separation in a Helmholtz double layer at the interface between the surface of a conducting electrode and an electrolyte, and pseudocapacitance – characterized by electrochemical storage of electrical energy. Pseudocapacitance occurs by redox reactions, electrosorption, or intercalation on the electrode surface by adsorbed ions which creates a reversible Faradaic charge transfer on the electrode. The contribution of double-layer capacitance or pseudocapacitance to supercapacitor operation depends on the electrode materials and structures (or microstructures) and the electrolyte composition. These features make supercapacitors similar to rechargeable batteries in the context of electrical energy

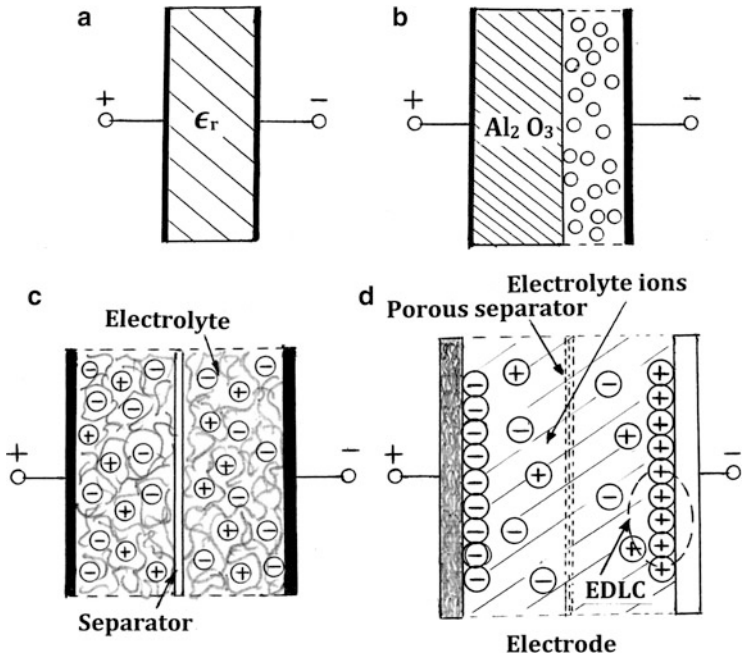


Fig. 29 Capacitor comparisons. (**a**) Dielectric (electrostatic) capacitor. (**b**) Electrolytic capacitor. (**c**) Electrochemical double] layer (supercapacitor). (**d**) Schematic showing details of two-cell electrode configuration for supercapacitor with electric double-layer capacitance (EDLC) and pseudocapacitance

storage or corresponding energy density devices. Figure 29 compares the principal design features for electrostatic, electrolytic, and electrochemical double-layer supercapacitors. It can be noted in Fig. 29c, d that the supercapacitor structure is similar to that of a rechargeable battery as illustrated generally in Figs. 26 and 28: consisting of two electrodes in contact with an electrolyte, with a separating membrane such as a porous polymer or ceramic. Like battery designs, supercapacitor electrodes are fabricated with porous carbon materials, especially nanomaterials: black carbons (BC) with nominal diameters of 30–50 nm (see Figs. 20 and 21). Of course, BC has often been and continues to be replaced with carbon nanotube arrays (CNTAs) or more efficient and cheaper graphene-based materials (Zhang et al. 2010; Liu et al. 2010; Tan and Lee 2013). However, graphene sheet structure in contrast to porous graphene structure is more efficient in supercapacitor operation at high frequencies because pores increase the device resistances at high frequency. Figure 30 shows these two forms of graphene. Graphene sheet structures as illustrated in Fig. 30b can be selectively spaced by aqueous electrolyte layers (such as KOH in water), although electrolyte compositions can also involve polyethylene oxide/$LiClO_4$ polymer electrolytes, conducting polyaniline (PANI), especially graphene/PANI composite structures as well as other graphene–polymer composite structures (Tan and Lee 2013).

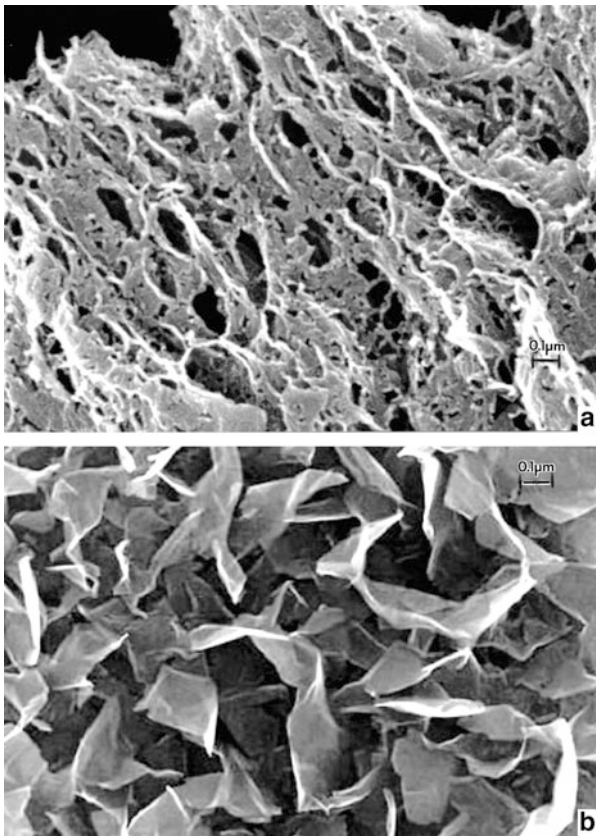


Fig. 30 Graphene structure (microstructures) observed in the SEM. (**a**) Porous graphene. (**b**) Nonporous, graphene nanosheets. (**a**) is courtesy of S. Kannappan, Gwangju Institute of Science and Technology, South Korea. (**b**) is courtesy of John R. Miller, JME Inc., Ohio

The energy storage (or specific energy) capability of a supercapacitor is represented by

$$E = CV^2/2 \tag{11}$$

where C is the capacitance and V is the cell voltage. The corresponding maximum power for a single-cell supercapacitor is given by

$$P = V^2/4\,R_s \tag{12}$$

where R_s is the equivalent series resistance of the cell. The capacitance, C, in Eq. 11 is dependent upon the electrode material, while the cell voltage is limited by the thermodynamic stability of the electrolyte. The electrode structure (such as porosity) and the electrolyte composition, ion mass transfer resistance, and contact resistance all contribute to R_s, which if kept small can have a significant effect on the power delivery as implicit in Eq. 12. As in rechargeable battery design, electrode surface area is a critical issue and is optimized by graphene technologies as in Li–air battery development as discussed above.

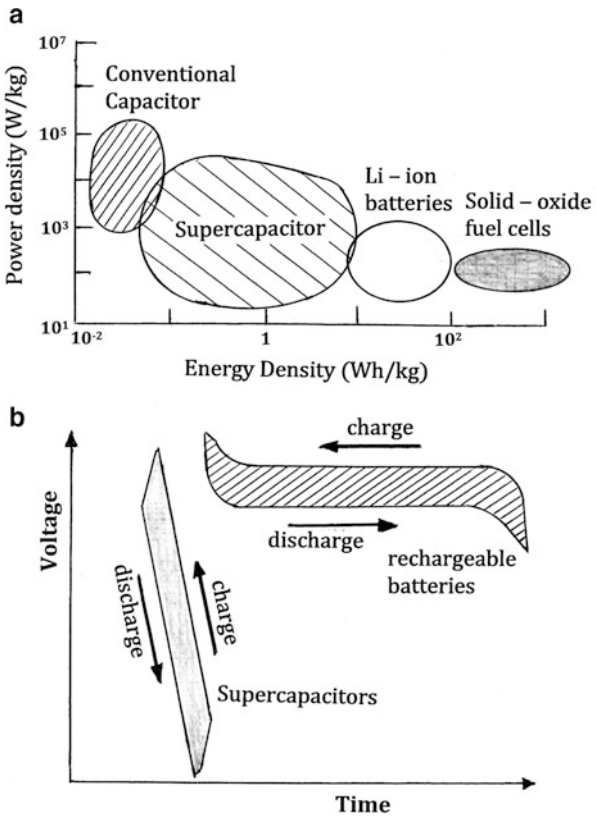


Fig. 31 Supercapacitor performance comparisons. (**a**) Ragone plot showing power density versus energy density for storage regimes. (**b**) Charge–discharge comparison for supercapacitors and rechargeable batteries (voltage versus time at constant current)

While the energy density of supercapacitors is much higher than conventional dielectric capacitors (Fig. 29a), it is lower than batteries and fuel cells. However, the power density can be more than 10 times greater than Li-ion rechargeable batteries, or solid-state fuel cells, as illustrated in Fig. 31a. In contrast to rechargeable batteries, supercapacitors exhibit considerably different and even complimentary charge–discharge behavior as shown in Fig. 31b. The implications of this behavior are that in electric vehicle or related power applications, supercapacitors can provide power bursts for start-up or restart, and power bursts required to adjust the blades of large wind turbines in changing wind conditions. However, with energy densities approaching 100 Wh/kg and charge times near 10 s, with essentially unlimited ($>10^5$) charge–recharge cycles, supercapacitors are beginning to compete for a wide range of rechargeable battery applications, including electric automobile requirements, and a wide range of portable electronic devices.

Like rechargeable battery systems and devices, solid-state, flexible, and transparent supercapacitor structures based on large-area graphene electrodes are being developed which will be easily manufactured and have a wide range of applications, especially substituting for rechargeable battery applications (Gao et al. 2013). These thin-film structures use poly(vinyl) alcohol (PVA)/phosphoric acid gel

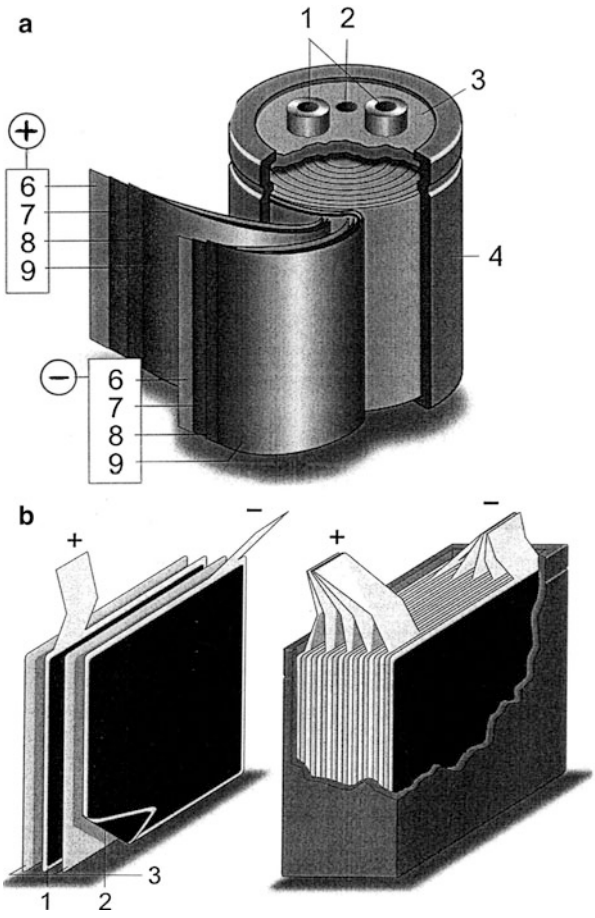


Fig. 32 Commercial supercapacitor (schematic) configurations. (**a**) Wound (cylindrical): *1* terminals, *2* vent, *3* seal, *4* Al can, *6* separator, *7* carbon electrode, *8* collector, *9* carbon electrode. (**b**) Stacked plate array. *1* positive electrode, *2* negative electrode, *3* separator (Wikimedia Commons images)

electrolytes between graphene electrodes, or fabricated on a polyethylene-terephthalate (PET) substrate to create a PET/PVA/H_3PO_4/graphene (thin-film) layer structure similar to that illustrated in Fig. 28b (Foa Torres et al. 2010).

The ability to create graphene layers by CVD as discussed previously in chapter "► Synthesis and Processing of Nanomaterials" can allow for the integration of micro-supercapacitors into a variety of electronic device structures, including ICs. Many sizes and configurations of supercapacitors are already commercially available as illustrated in Fig. 32.

Fuel Cells

Fuel cells are devices which convert energy from a fuel into electricity by reaction with oxygen. Fundamentally, a fuel cell design resembles a Li–air battery (Fig. 28) with the Li anode being replaced by a catalyst-containing electrode. Hydrogen is

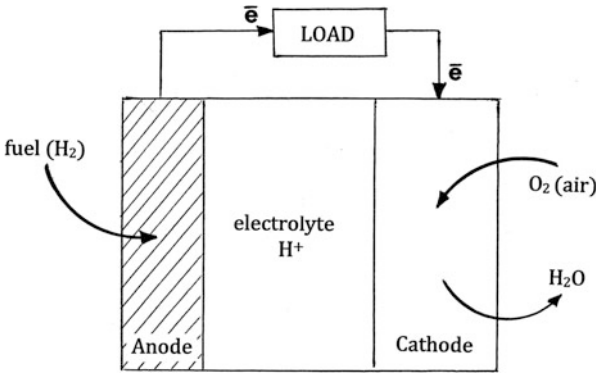


Fig. 33 Simple fuel cell schematic. The electrolyte supports proton (H^+) conduction

the preferred fuel which combines with oxygen at the anode to produce water and electrons, while at the cathode oxygen plus the electrons produced at the anode yield water:

$$\begin{aligned} &\text{Anode Reaction}: 2H_2 + 2O^{2-} \rightarrow 2H_2O + 4\bar{e} \\ &\text{Cathode Reaction}: O_2 + 4\bar{e} \rightarrow 2O^{2-} \end{aligned} \tag{13}$$

Figure 33 shows a simple fuel cell schematic diagram. As in rechargeable battery designs and supercapacitors, fuel cells employ an electrolyte that allows charges (H^+ in Fig. 33) to move between the two fuel cell electrodes; electrons are drawn from the anode to the cathode through an external circuit or load (Fig. 33). In addition to hydrogen, fuels can include most any combustible hydrocarbon. While many fuel cell electrolytes are the same as those in rechargeable batteries and supercapacitors, e.g., aqueous alkaline solutions (KOH) and molten phosphoric acid (H_3PO_4), solid oxide fuel cells (SOFCs) use a ceramic material like yttria-stabilized zirconia (YSZ) as the electrolyte, although these fuel cells require hydrogen fuel or a fuel converted to pure hydrogen.

Like other electrochemical systems (batteries and supercapacitors), fuel cell electrodes are fabricated from carbon structures, including graphene having microstructures illustrated in Fig. 30, intercalated with Pt, Pd, or Co nanoparticle catalysts. In SOFC cells, anode structures have been porous Ni-YSZ, while the cathodes have been composed of porous lithium strontium manganite (LSM) (Haile 2003). Jean et al. (2013) have recently shown that graphene nanoplatelets similar to those shown in Fig. 30b outperform metal catalyst electrodes.

In contrast to rechargeable battery and supercapacitor technologies, fuel cell technologies, especially hydrogen fuel cell technologies, lag considerably behind because of the difficulty in developing a hydrogen fuel infrastructure in contrast to charging infrastructure required for battery and supercapacitor charging technologies. However, like rechargeable battery and supercapacitor thin-film technologies, thin-film heterostructure solid oxide fuel cells have been designed based on microlithography and related thin-film deposition processes (Chen et al. 2004; Winter and Brodd 2004).

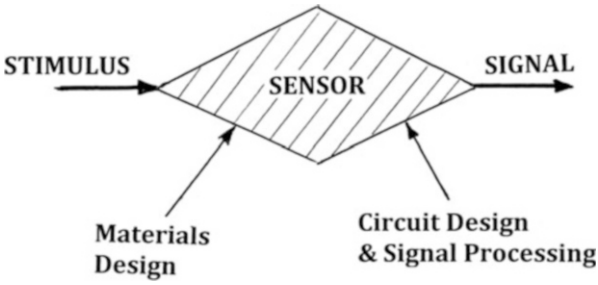


Fig. 34 Simple sensor features

Sensor Materials and Devices

The fundamental principles involved in the ability of materials or device configurations to sense or measure a property often rely upon resistance, capacitance, or inductance variations which depend upon displacements over length scales, adsorption and reaction of species dependent upon specific surface area, and intrinsic electronic structures, especially bandgap energy. Consequently materials selection and design, including manipulation of structure and microstructure, and electrical circuit design and signal processing apply generally to sensor development as illustrated schematically in Fig. 34. Sensors and sensor device development epitomize the materials structure – properties–processing–performance paradigm (Byrnes and Ostheimer 2006; Fraden 2010; Koroteenkov 2014). Thin-film layer configurations which can be fabricated in place or easily attached to surfaces for monitoring internal or external responses are particularly important in a wide range of automotive, aerospace, and energy-related application areas. Strain, temperature, pressure, humidity, radiation, gas, and other chemical species detection, measurement, and monitoring are especially prominent features of sensor development and application.

Metal oxides and other wide-bandgap semiconducting materials are utilized in many solid-state detector configurations ranging from radiation to gas detection. Simple sandwich configurations of metal electrode/semiconductor detector/metal electrode (MSM) devices illustrated conceptually in Fig. 35a are especially useful because of their ability to be fabricated using IC technologies or 3D printing concepts. Targeted selectivities have involved Fe_2O_3 for CO detection; SnO_2 for H_2, CO, NO_2, CH_4, and H_2S; TiO_2 for H_2; C_2H_5OH (alcohol), O_2, and ZnO for CH_4; and Al_2O_3 for humidity (H_2O) sensing along with conducting nanometals such as Ag.

Oxide films such as TiO_2 loaded with nanometal particles such as Au or Ag can alter the effective bandgap energy and change the adsorption of gas species or absorption of radiation, especially in the visible portion of the spectrum. In addition, oxygen vacancies on metal-oxide surfaces are electrically and chemically active since they can function as *n*-type donors and can change the conductivity, σ. Adsorption of charge-accepting molecules such as NO_2 or O_2 at vacancy sites reduces σ by electron depletion from the conduction band, while molecules such

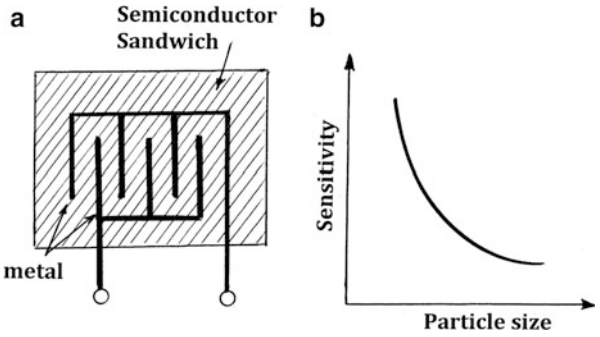


Fig. 35 (**a**) Interdigitated MSM sensor device schematic. (**b**) Idealized device sensitivity versus particle size

as H_2 or CO react with surface adsorbed O and remove it, causing σ to increase. Adsorption of molecules of various types in channel arrays forming electrically continuous structures can change the conductivity as well. Such channel arrays can be manipulated by growing tubes or rods of conducting or semiconducting material where molecular adsorption or the quantity of adsorbed molecules can be correlated with changes in conductivity. In many configurations involving nanoparticles or nanostructures, the particle size or nanodimension as it relates to specific surface area usually influences sensitivity as shown generally in Fig. 35b, often referred to as an interdigitated sensor device.

In addition to particle size, sensitivity as shown ideally in Fig. 35b is similarly influenced by grain size in fabricated thin-film arrays, in some cases because the grain boundary influences adsorption of species different from the grain matrix as a consequence of energetics, and of course the conductivity will intrinsically be influenced by grain size or grain boundary (interfacial) area. Of course it is possible to fabricate multilayer, hybrid, or multifunctional thin-film sensor arrays which can simultaneously sense numerous properties or chemistries simultaneously. This represents an extensive regime for materials innovations (Zribi and Fortini 2009).

References

Araujo GL, Marti A (1994) Absolute limiting efficiencies for photovoltaic energy conversion. Solar Energy Mater Solar Cells 90:1068–1075

Besenhard JO (ed) (1999) Handbook of battery materials. Wiley-VCH, New York

Brock DC (ed) (2006) Understanding Moore's law: four decades of innovation. Chemical Heritage Press, Philadelphia

Burschka J, Pellet N, Moon S-J, Humphrey-Baker R, Gao P, Nazeeruddin MK, Grätzel M (2013) Sequential deposition as a route to high-performance perovskite-sensitized solar cells. Nature 499:316–319

Byrnes J, Ostheimer G (eds) (2006) Advanced sensing with security applications, vol 2, Nato security through science series. Springer, New York

Cai JW, Mahan GD (2008) Transport properties of quantum dot arrays. Phys Rev B 78:035115–035118

Castellano JA (2005) Liquid Gold: The story of liquid crystal displays and the creation of an industry, World Scientific, New Jersey

Chandrasekhar S (1992) Liquid crystals. Cambridge University Press, Cambridge, UK
Chen X, Wu NJ, Smith L, Ignatiev A (2004) Thin-film heterostructure solid oxide fuel cells. Appl Phys Lett 84(14):2700–2702
Chen J, Cranton W, Fihn M (eds) (2012) Handbook of visual display technology. Springer, New York
Dragoman D, Dragoman M (2007) Giant thermoelectric effect in graphene. Appl Phys Lett 91(20):203116–203119
Foa Torres LEF, Roche S, Charlier J-C (2014) Introduction to graphene-based nanomaterials: from electronic structure to quantum transport. Cambridge University Press, Cambridge, UK
Fraden J (2010) Handbook of modern sensors. Springer, New York
Friedman DJ, Olson JM, Kurtz S (2011) High efficiency III–V multijunction solar cells. In: Luque A, Hegedue S (eds) Handbook of photovoltaic science and engineering, 2nd edn. Wiley, New York, pp 314–358
Gao Y, Zhou YS, Xiang W, Jiang LJ, Mahjouri-vamani M, Thirugnanam P, Huang X, Wang MM, Jiang L, Lu YF (2013) Transparent flexible, and solid-state supercapacitors based on graphene electrodes. APL Mater 1:012101-1–012101-7
Goldsmid HJ (1964) Thermoelectric refrigeration. Plenum Publishing, New York
Grätzel M (2003) Dye-sensitized solar cells. J Photochem Photobiol C Photochem Rev 4(2):145–167
Gross L, Mohn F, Moll N, Liljeroth P, Meyer G (2009) The chemical structure of a molecule resolved by atomic force microscopy. Science 325(5944):1110–1114
Haile SM (2003) Fuel cell materials and components. Acta Mater 51:5981–6000
Heath JR (2009) Molecular electronics. Ann Rev Mater Res 39:1–23
Hicks LD, Dresselhaus MS (1993a) Effect of quantum-well structures on the thermoelectric figure of merit. Phys Rev B 47(19):12727–12730
Hicks LD, Dresselhaus MS (1993b) Thermoelectric figure of merit of a one-dimensional conductor. Phys Rev B 47(24):16631–16634
Jean IY, Choi HJ, Choi M, Seo JM, Jung SM, Kim MJ, Zhang S, Zhang L, Xia Z, Dai L, Park N, Back JB (2013) Facile, scalable synthesis of edge-halogenated graphene nanoplatelets as efficient metal-free electrocatalysts for oxygen reduction reaction. Nature scientific reports 3, article no. 1810
Koroteenkov G (2014) Handbook of gas sensor materials. Springer, New York
Koumoto K, Wang Y, Zhang R, Kosuga A, Funahashi R (2010) Oxide thermoelectric materials: a nanostructuring approach. Ann Rev Mater Res 40:363–394
Kumar B, Kumar J, Leese R, Fellner JP, Rodrigues SJ, Abraham KM (2010) A solid-state rechargeable, long cycle life lithium-air battery. J Electrochem Soc 157:A50–A65
Landsberg PT, Tonge G (1980) Thermodynamic energy conversion efficiencies. J Appl Phys 51(7):R1–R20
Li J-F, Liu W-S, Zhao L-D, Zhou M (2010) High-performance nanostructured thermoelectric materials. NPG Asia Mater 2(4):152–158
Li Y, Gong M, Liang Y, Feng J, Kim JE, Wang H, Hong G, Zhang B, Dai H (2013) Advanced zinc-air batteries based on high-performance hybrid electrocatalysts. Nat Commun 4:1805–1808
Li XA, Ren RY, Ren MW, Ma YW, Yang JP, Bao G (2013) Integration of graphene into thin film transistors. Mater Technol 28(4):175–180
Liu C, Yu Z, Neff D, Zhamu A, Jang BZ (2010) Graphene-based supercapacitor with an ultrahigh energy density. Nano Lett 10:4863–4868
Luque A (2011) Will we exceed 50% efficiency in photovoltaics. Appl Phys Rev 110:031301-1–031301-19
Marti A, Araujo GL (1996) Limiting efficiencies for solar conversion in multigap systems. Solar Energy Mater Solar Cells 43(2):203–222
Micheloni R, Crippa L, Marelli A (2010) Inside NAND flash memories. Springer, New York
Micheloni R, Marelli A, Eshghi K (2013) Inside solid state drives (SSDs). Springer, New York
Murr LE (1978) Solid-state electronics. Marcel Dekker, New York

Nair RR, Blake P, Grigovenko AN, Navoselov KS, Booth TJ, Stauber T, Peres NMR, Geim AK (2008) Fine structure constant defines visual transparency of graphene. Science 320(5881): 1308–1311

O'Regan B, Grätzel M (1991) A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films. Nature 353(6346):737–740

Park K-T, Shin S-M, Tazebay AS, Um H-D, Jung J-Y, Jee S-W, Oh M-W, Park S-D, Yoo B, Yu C, Lee J-H (2013) Lossless hybridization between photovoltaic and thermoelectric devices. Sci Rep 3(2123):1–6

Pichanusakorn P, Bandaru P (2010) Nanostructured thermoelectrics. Mater Sci Eng R 67:19–63

Pikus FF, Likharev KK (1997) Nanoscale field-effect transistors: an ultimate size analysis. Appl Phys Lett 71(25):3661–3664

Rowe DM (ed) (2006) Thermoelectric handbook: macro to nano. CRC/Taylor & Francis, Boca Raton

Sala G, Anton I (2011) Handbook of photovoltaic science and engineering, 2nd edn. Wiley, New York

Senevirathne K, Day CS, Gross MD, Lachgar A, Holzwarth NAW (2013) A new crystalline LiPON electrolyte: synthesis, properties and electronic structure. Solid State Ion 233(21): 95–101

Shakouri A (2011) Recent developments in semiconductor thermoelectric physics and materials. Ann Rev Mater Res 41:399–431

Shockley W, Queissen HJ (1961) Detailed balance limit of efficiency of p-n junction solar cells. J Appl Phys 32:510–514

Tan YB, Lee J-M (2013) Graphene for supercapacitor applications. J Mater Chem A 1:14814–14843

Tributsch H, Calvin M (1971) Electrochemistry of excited molecules: photoelectrochemical reactions of chlorophylls. J Photohem Photobiol 14(14):95–112

Winter M, Brodd J (2004) What are batteries, fuel cells and supercapacitors? Chem Rev 104:4245–4265

Zhang X, Chau KT, Chan CC (2009) Design and implementation of a thermoelectric-photovoltaic hybrid energy source for hybrid electric vehicles. World Electric Vehicle J 3:1–11

Zhang LL, Zhou R, Zhao XS (2010) Graphene-based materials as supercapacitor electrodes. J Mater Chem 20:5983–5992

Zribi A, Fortini J (2009) Functional thin films and nanostructures for sensors. Springer, New York

Photonic Materials and Structures

Contents

Abstract

In contrast to conventional or even natural materials which derive their electromagnetic characteristics from the properties of the composing atoms and molecules, photonic crystals and related metamaterials allow for the design of meta-atoms or superatoms as well as artificial structures which allow for new optical and electromagnetic functionalities. These features are briefly presented in this chapter beginning with tunable photonic crystal structures to fabricate an array of inks activated by electric and magnetic fields as well as external stresses. The manipulation of permittivity in dielectric materials and permeability, especially in developing composite structures with negative index of refraction, is discussed in the context of fabrication permitted by 3D printing and related technologies.

Introduction

Photonics relates generally to manipulation of light and the particle properties of light. This includes the generation and emission of light, as well as its propagation, modulation, and amplification in materials or material systems, and photonic devices which incorporate photonic crystals and related photonic structures, including metamaterials, with engineered optical properties. These structures are ubiquitous in nature where structural colors are observed in opals and opal-like and related

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_59

photonic crystal structures in many living species, with numerous examples already illustrated in ▶ Fig. 14 in chapter "Electromagnetic Color and Color in Materials" and ▶ Figs. 7 and ▶ 8 in chapter "Biomimetics and Biologically Inspired Materials," especially structural colors.

Figure 1 illustrates the periodic structures which characterize photonic crystals (Joannoupoulas et al. 1995), a concept for a class of materials first proposed independently by S. John and E. Yablonovitch around 1987 (John 1987; Yablonovitch 1987). One of the fundamental features of 3D photonic crystals is the existence of an optical bandgap which can influence photon propagation and behavior in an analogous way that an electronic bandgap in regular crystal structures, especially semiconducting crystal structures, can influence electron propagation. In principle, the 3D photonic crystal band structure can be obtained by solving Maxwell's equations, with the dielectric constant expressed as a 3D periodic function. Yablonovitch (1987) used a scalar model to argue that a face-centered cubic lattice should be favored for the formation of a complete bandgap although more rigorous models were necessary to account for the tensor properties of electromagnetic wave components (e.g., **B** and **D**; see chapter "Electromagnetic Fundamentals"). It is generally accepted that for an fcc lattice consisting of spherical building blocks or superatoms, there only exists a pseudo gap in the photonic band structure regardless of the contrast in dielectric constant.

Figure 2 illustrates the 3D photonic fcc crystal structure implicit in natural opal, where variations in the size of SiO_2 superatoms in the (111) close-packed plane give rise to structural color variations or reflected light wavelength variations implicit in the Bragg-Snell equation:

$$\lambda = 2D\left(n_{eff}{}^2 - \cos^2\theta\right)^{1/2} \tag{1}$$

which is characteristic for the geometrical optics for incident light on the (111) crystal plane on an opal (Fig. 2b) as shown schematically in Fig. 3. In Eq. 1, D is the distance between the (111) opal (or photonic crystal) lattice planes (or interplanar spacing): $D = \sqrt{2/3}\,d$, where d is the sphere or superatom diameter and θ is the (Bragg) angle of incidence of light falling on the opal. The volume-weighted average of the refractive index or effective refractive index is given by

$$n_{eff} = 10.74\, n_{sphere} + 0.26\, n_{void}) \tag{2}$$

where n_{sphere} for the natural opal in Fig. 2b represents the index of refraction for SiO_2 (1.54) while n_{void} is the air space between the SiO_2 spheres ($n = 1$). However, for fabricated 3D fcc photonic crystals, the superatom spheres can be made of a variety of materials and the space between them can also be another material or dielectric, having $n > 1$. In this case, the effective refractive index becomes

$$n_{eff} = (0.74\, n_s + 0.26\, n_{im}) \tag{3}$$

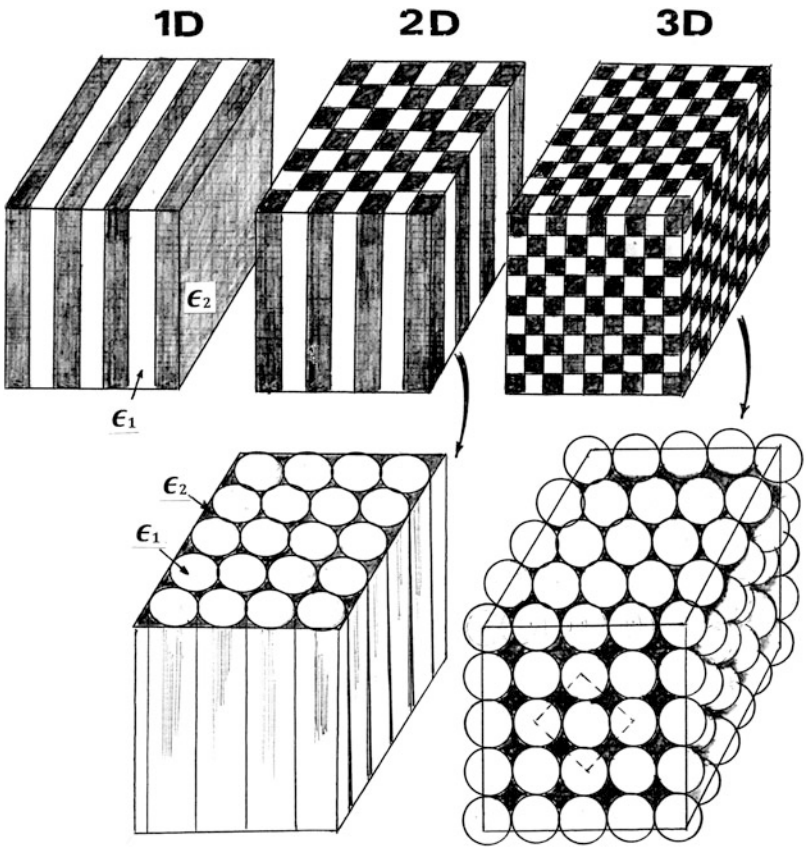


Fig. 1 Schematic illustrations of 1D, 2D, and 3D photonic crystal structures composed of two different dielectric materials. Permittivities are designated ϵ_1, ϵ_2, $\epsilon_{r1} = \epsilon_1/\epsilon_0$; $\epsilon_{r2} = \epsilon_2/\epsilon_0$. Note the 2D cylindrical and 3D sphere structures, where the *dotted square* in the latter represents the fcc crystal structure unit

where n_s is the sphere or superatom refractive index and n_{im} is the refractive index of the interstitial material or the dispersant matrix as implicit in Fig. 1. Ideally, it is possible to create an inverse opal by fabricating a 3D photonic crystal structure where the spheres are etched away, leaving the matrix as a solid. This simply alters the effective index of refraction implicit in Eq. 3.

It can be observed in Eq. 1 that color tuning of a photonic crystalline material (as implicit in the many opal colors shown in Fig. 2a) can be achieved by changing or manipulating the sphere size, d, and correspondingly D, as well as the refractive index of the spheres or sphere material, and/or its dispersant or matrix refractive index (Eq. 3), creating active opal color. This can be applied to either 3D or 2D photonic crystal structures as shown in Fig. 1.

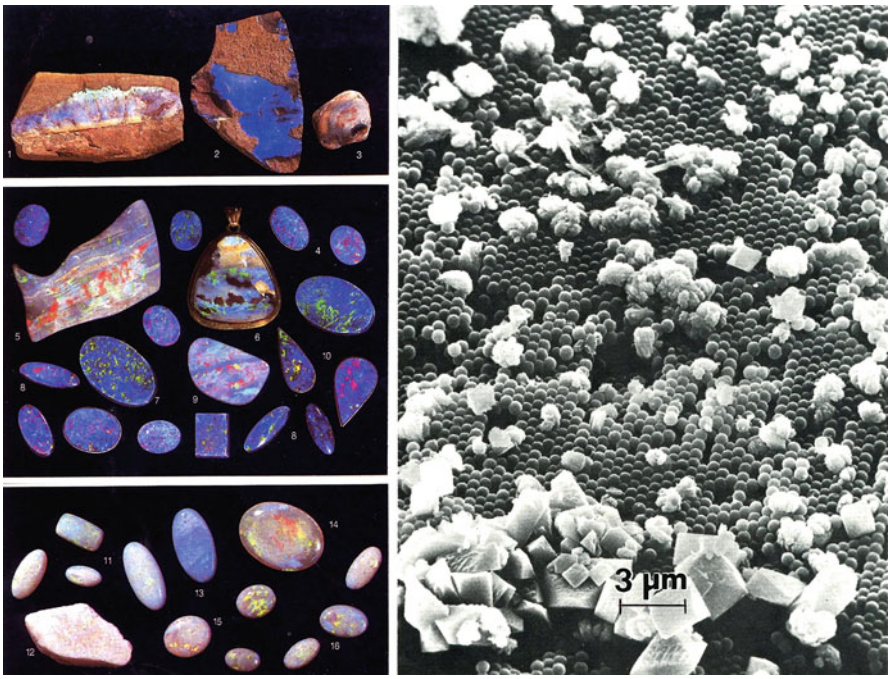


Fig. 2 Natural opals (**a**) and opal structure (**b**). (**a**) Opal color variations arise by variations in SiO_2 sphere size in photonic crystal lattice as shown in (**b**). Pyramidal crystals are β-SiO_2 (cristobalite). SiO_2 spheres are in (111) fcc close-packed arrays

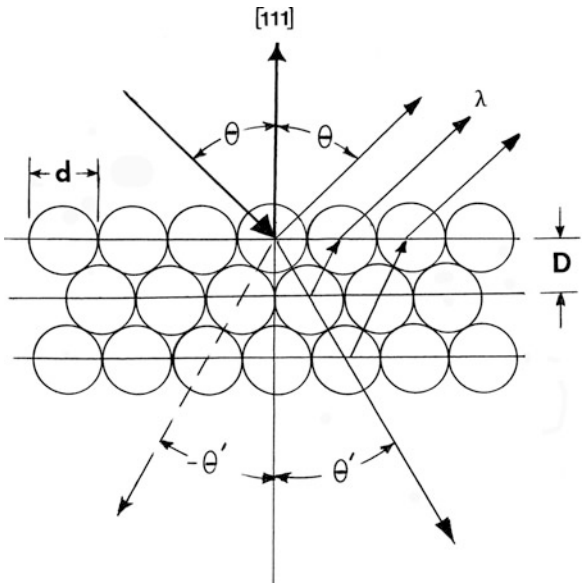


Fig. 3 Reflection and diffraction from (111) crystal plane in opal. Refractive angle is denoted θ′

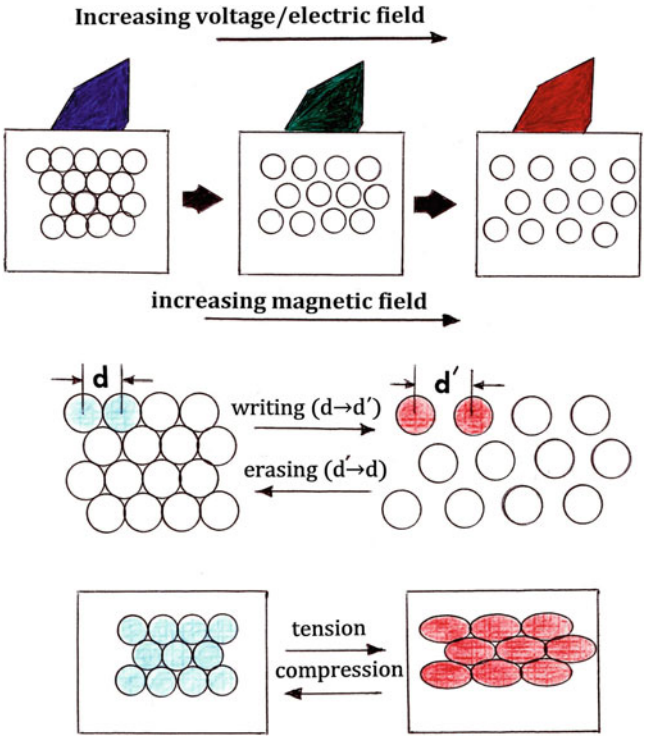


Fig. 4 Tunable photonic inks. (**a**) Ceramic or polymer spheres in polymer matrix for electric displacement or anisotropic expansion or magnetic spheres in polymer matrix for magnetic displacement. (**b**) Dielectric or magnetic spheres in solvent on paper for photonic write-erase concept. (**c**) Compressive or tensile (stretching) strain of elastic opal or inverse opal (matrix) structure

Electrically, Magnetically, and Mechanically Tunable Opals: Photonic Inks and Color Display Systems

Within the past decade, novel applications of Eq. 1 in this general form have allowed for the development of a variety of photonic inks or *P-inks* comprised of an opal or superatoms embedded in a matrix or dispersant to be spatially manipulated in an electric or magnetic field as illustrated schematically in Fig. 4a. In the case of electrically tuned P-ink, the applied voltage or electric field causes the dispersant (often an engineered gel) to expand or shrink, changing the sphere spacing, d to d′ or d′ to d. In the case of magnetic field tuning, the spheres are magnetic colloidal (nano) particle spheres such as Fe_3O_4 (magnetite) in a flexible polymer matrix, in which case the magnetic field displaces the magnetic particles and not the matrix. Wang et al. (2013) have more recently described a magnetically rewritable photonic ink *(M-ink)* based on superparamagnetic nanochains. Electrically tunable photonic inks

or P-inks have consisted of polystyrene nanospheres in an elastomeric matrix such as polydimethylsiloxane (PDMS) or aliphatic compound dispersant solvents which can be spread on various composition papers to write or erase as shown schematically in Fig. 4b. SiO_2 nanospheres in a polyferrocenylsilane (a metallopolymer network) matrix is also a viable P-ink. Ge et al. (2007) have also described polymeric opals consisting of a hard polystyrene core coated with a thin polymer layer containing allyl methacrylate and a soft polyethyl acrylate outer shell which could be prepared as fibers and woven into fabrics which could change color when stretched (Fig. 4c). Indeed, as shown in Fig. 4c, a stretchable surface (polymer or metal) coated with an opal photonic crystal film or *elast-ink* can change color. Such a color response to strain could of course serve as a strain sensor or strain gauge.

By coating opal-based photonic inks onto an array of pixels, a full color display can be created as originally envisioned by Ozin and Arsenault (2008); wherein the color of each pixel can be varied. In addition, since this display is reflective, it is not overwhelmed by glare in sunlight as contemporary displays such as LCD or LED displays. In addition, photonic crystal-based microarray applications can be fabricated by high-density inkjet (3D) printing as recently discussed by Lai et al. (2011).

Photonic crystal coatings can also be applied to a surface through emersion in a dilute colloidal suspension of superatoms or meta-atoms. Often referred to as vertical evaporation-induced self-assembly, it consists of submersion of a flat substrate in a dilute colloidal suspension where sphere deposition (and self-assembly) occurs in a crystalline (close-packed) arrangement at the meniscus as the solvent evaporates. The layers can be manipulated by spin coating, for example.

Photonic Metamaterials and Nanophotonic Device Concepts

It is perhaps implicit in Fig. 4 that photonic printing strategies involving the self-assembly of colloidal crystals locally on a substrate and the change of photonic structure by adjusting or tuning the lattice spacing, d; crystal orientation, θ (Fig. 3); the permittivity, ϵ; or the index of refraction, n, including composite crystals (Fig. 1), can exhibit *invisible photonic printing*, particularly useful for steganography and watermarking for anticounterfeiting and other novel applications (Fudouzi and Xia 2003; Dolling et al. 2006; Xuan and Ge 2011, 2012; Hu et al. 2013). In addition, and as implicit in Fig. 1, the ability to create negative refraction ($-\theta'$) or a corresponding negative index of refraction ($-$n),

$$N = -\sqrt{\varepsilon\mu}, \tag{4}$$

where both ε and μ are negative ($\varepsilon < 0$; $\mu < 0$) as originally discussed by Veselago (1968), can allow for a wide range of unconventional and unnatural optical phenomena, including invisibility in the visible portion of the spectrum as discussed previously in Part III and chapter "▶ Novel Structure Printing" (Fig. 2).

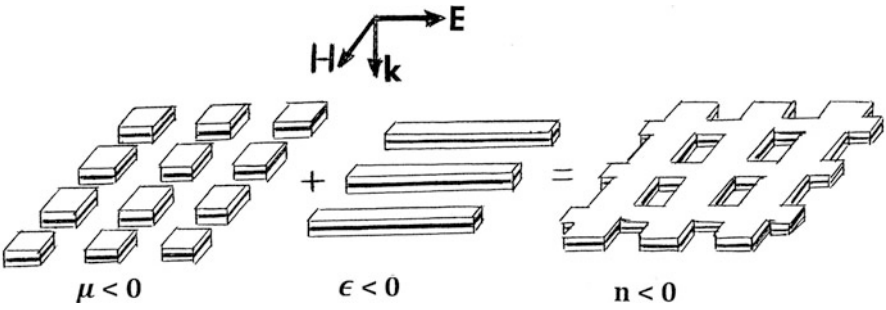


Fig. 5 Conceptual negative-index metamaterial design adapted from Dolling et al. (2006). Incident electric and magnetic field vectors are denoted **E** and **H**, respectively. **K** represents the incident light wave vector. Note that the size and layer structure are variables

Such unusual photonic metamaterials can be envisioned as a composite composed of photonic metamaterial blocks corresponding to $\mu < 0$ and composite wires or rods corresponding to $\varepsilon < 0$ which when combined into a photonic metamaterial composite produces a structure where $n < 0$. Such novel composite structures can be ideally fabricated using combinations of metals, especially good conductors like Au or Cu, and dielectric (or insulating) materials. Such structures interact more strongly with magnetic fields than conventional or natural materials, particularly at frequencies (or wavelengths) that have resonances with the structures. Consequently, varying the shape of the structure or the configuration of the building blocks as implicit in Fig. 5 or conceptually as shown in ▶ Fig. 2 (chapter "Novel Structure Printing") can create gradients in optical properties that can manipulate light: transformation optics, where structures transform the wavefront of light (Wegener and Linden 2010). The use of nanostructured materials with very large permittivity offers the prospect for manipulating the electric and magnetic response using only dielectric metamaterials. To date, it has not been possible to fabricate cloaking materials for the visible portion of the spectrum, however. On the other hand, novel resonance structures, particularly 3D cloaking structures, may be developed using aperiodic or photonic quasicrystal structures which are possible to fabricate using stereolithography (Part XI) as illustrated conceptually in Fig. 6 (see chapter "▶ Aperiodic Crystal Structures: Quasicrystals"). Vardeny et al. (2013) have recently reviewed optical transmission and reflectivity in one-, two-, and three-dimensional photonic quasicrystals.

In addition to the many novel innovations in photonic metamaterial structure development and fabrication, the fabrication and application of a number of photonic structures such as photonic quantum dot arrays, multi-quantum well (MQW) arrays, and integrated photonic circuits (IPCs) are examples of photonic materials and device innovations (Vivien and Pavesi 2013). The hybridization of electronics with photonics in particular is a growing technological area.

High-brightness LED configurations using nanorod arrays, especially GaN nanocolumns or complex, single-crystal, columnar fibers such as InGaN/GaN or InGaN multiple quantum disks (MQDs) embedded in GaN nanocolumns grown by

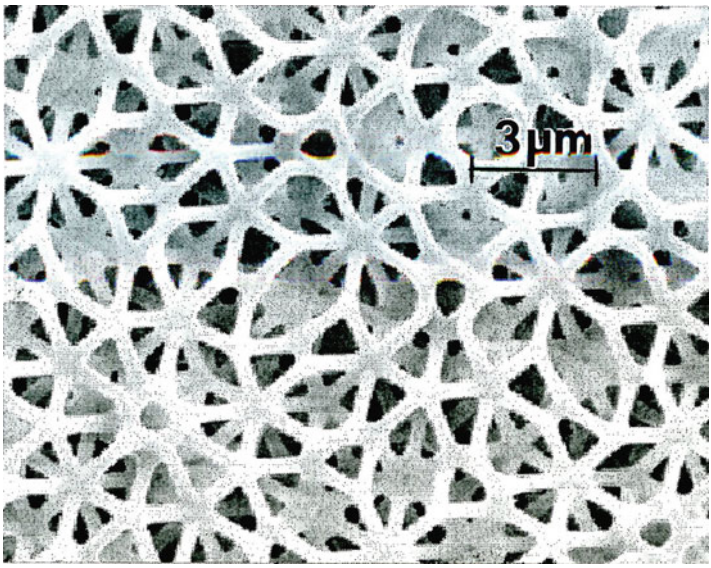


Fig. 6 Fivefold symmetric 3D photonic quasicrystal structure (Adapted from Ledermann et al. (2006))

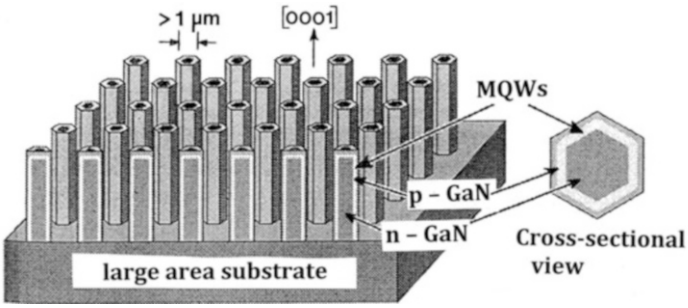


Fig. 7 GaN-based nano LED schematic. The nanorod structure can be InGaN/GaN core-shell structure which grows naturally (Adapted from Li and Wang (2012)). Multiple quantum wells (MQWs)

MBE on Si, or other semiconductor substrates function as efficient LEDs (Liao et al. 2012; Li and Wang 2012). Figure 7 illustrates these concepts schematically.

It is interesting to note that while quantum dots having sizes < 6 nm provide electron confinement, Bayer et al. (1998) demonstrated that photons can also be confined within what was referred to as a photonic molecule wherein the confined photons are restricted to a set of discrete energies, like electrons in real molecules. To achieve this, two semiconductor GaAs blocks 3 μm on a side and 1 μm in height were linked together with a narrow bridge to allow them to interact. A tiny quantum

well situated in the center of these blocks produced light when pulsed by a laser, and as the bridge spacing was shortened, the photonic atom states transformed into photonic molecule states. Such coherent interactions between individual photons, like those associated with particles having measurable mass, could have a variety of novel scientific and engineering applications. In fact, Firstenberg et al. (2013) more recently demonstrated a quantum nonlinear medium inside which individual photons propagated as massive particles with strong mutual attraction: a two-photon-bound state creating a photon molecule. This raises the prospect for realizing a single-photon transistor recently described by Chen et al. (2013).

References

Bayer M, Gutbrod T, Reithmaier JP, Forchel A, Reinecke TL, Knipp PA, Dremin AA, Kulakovki VD (1998) Optical modes in photonic molecules. Phys Rev Lett 81: 2582–2585

Chen W, Beck KM, Bücker R, Gullans M, Lukin MD, Tanji-Suzuki H, Vuletic V (2013) All-optical switch and transistor gated by one stored photon. Science 341(6147):768–770

Dolling G, Enkrick C, Wegener M, Soukoulis CM, Linden S (2006) Simultaneous negative phase and group velocity of light in a metamaterial. Science 312:892–894

Firstenberg O, Peyronel T, Liang Q-Y, Gonshkov AV, Lukin MD, Vuletic V (2013) Attractive photons in a quantum nonlinear medium. Nature 502:71–75

Fudouzi H, Xia YN (2003) Photonic papers and inks: color writing with colorless materials. Adv Mater 15:892–896

Ge J, Hu Y, Yin Y (2007) Highly tunable superparamagnetic colloidal photonic crystals. Angew Chem Int Ed 46(39):7428–7431

Hu H, Tong J, Zhong H, Xi Z, Chen C, Chen Q (2013) Invisible photonic printing: computer designing graphics, U.V. printing and shown by a magnetic field. Sci Rep (Nature) 3:1484 (1–5)

Joannoupoulos JD, Meade RD, Winn JN (1995) Photonic crystals. Princeton University Press, Princeton

John S (1987) Strong localization of photons in certain disordered dielectric superlattices. Phys Rev Lett 58(23):2486–2489

Lai W-C, Moncivais K, Chakravarty S, Wang X, Lin C-Y, Zhang ZJ, Chen RT (2011) High density inkjet printing of bio-molecules for photonic crystal-based microarray applications. In: Biochemical sensor conference proceedings, 12–15 June 2011

Ledermann A, Cademartiri L, Hermatschweiler M, Torninelli C, Orzin GA, Wiersma DS, Wegener M, Von Freymann G (2006) Three-dimensional silicon-inverse photonic quasicrystals for infrared wavelengths. Nat Mater 5:942–945

Li S, Wang A (2012) GaN based nanorods for solid state lighting. J Appl Phys 111:071101-1-15

Liao C-H, Chang W-M, Chen H-S, Chen C-Y, Yao Y-F, Chen H-T, Su C-Y, Ting S-Y, Kiang Y-W, Yang CC (2012) Geometry and composition comparisons between c-plane disc-like and m-plane core-shell InGaN/GaN quantum wells in a nitride nanorod. Opt Express 20(14): 15859–15864

Ozin GA, Arsenault AC (2008) P-ink and Elast-ink from lab to market. Mater Today 11(7–8): 44–51

Veselago VG (1968) The electrodynamics of substances with simultaneously negative values of ε and μ. Sov Phys Usp 10:509–514

Vivien L, Pavesi L (eds) (2013) Handbook of silicon photonics. CRC Press, Boca Raton

Vardeny ZV, Nahata A, Agrawal A (2013) Optics of photonic quasicrystals. Nat Photonics 7:177–187

Wang M, He L, Hu Y, Yiu Y (2013) Magnetically rewritable photonic ink based on superparamagnetic nanochains. J Mater Chem C 1:6151–6156
Wegener M, Linden S (2010) Shaping optical space with metamaterials. Phys Today 63:L32–L36
Xuan RY, Ge PJ (2011) Photonic printing through the orientational tuning of photonic structures and its application to anticounterfeiting labels. Langmuir 27:5694–5699
Xuan RY, Ge PJ (2012) Invisible photonic prints shown by water. J Mater Chem 22:367–372
Yablonovitch E (1987) Inhibited spontaneous emission in solid-state physics and electronics. Phys Rev Lett 58(20):2059–2062

Part XVI

Computational Modeling and Simulation of Materials and Materials Systems

Computer Simulation in Materials Science and Engineering

Definitions, Types, Methods, Implementation, Verification, and Validation

Contents

Abstract

Computer simulation is described as a comprehensive method for studying materials and materials systems. Computational methods used on different length and time scales for the simulation of materials structures and behavior are described along with process features involved in the implementation, verification, and validation of computer simulations. Computer simulation in the context of integrated computational materials engineering as this relates to the materials genome initiative concept for materials innovation and advanced materials development and deployment is discussed.

Introduction

Computational physics, computational chemistry, computational materials science and chemistry, and computational materials science and engineering embody computational methods used on different length and time scales for the simulation of materials behavior. These involve the integration of materials synthesis, processing, characterization, theory, and experiments along with computer simulation and modeling. Verification and validation of various computational approaches and simulations that span the wide ranges in time and length scales is also an important aspect of computer simulations in materials sciences and engineering. A growing extension of computational materials science and engineering also involves the integration of materials information or informatics with product

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_60

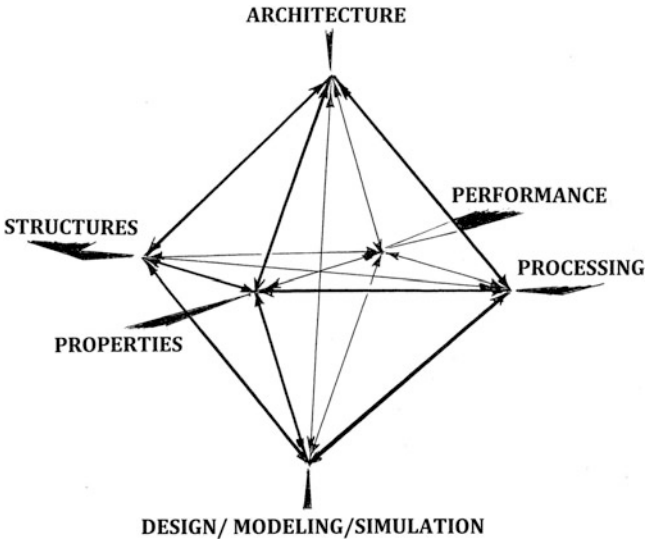


Fig. 1 The extended octahedron for materials science and engineering. See Fig. 24 in chapter "▶ Laser and Electron Beam Melting Technologies"

manufacturing process simulation and performance analysis, which in effect extends or reconfigures the materials science and engineering tetrahedral or pyramidal paradigm to one involving an octahedron as illustrated in Fig. 1. Often referred to as integrated computational materials engineering or integrated computational materials science and engineering, this multiscale modeling strategy employs computer simulations in shortening design and manufacturing cost and time as well as the implementation of new materials and manufacturing technologies.

Computational materials science and engineering, to be effective, needs to be driven by new insights and theories vested in engineering materials performance along with the other paradigm elements shown in Fig. 1. In addition, computational modeling should be routinely integrated as far as possible with experimentation and the implementation of the new materials paradigm, including characterization, which can provide real images and microstructures across the length scale in particular.

In the broadest sense, computer simulation can be used for prediction, qualitative or systemic predictions, and to understand materials systems and their behavior. Predictive capability drives technological discovery and innovation in materials science and engineering. Where data is available for a materials system, computer simulation can determine how events could have occurred or how events actually did occur. Broadly speaking, computer simulation is a comprehensive method or process for studying materials systems. This process includes choosing a model and methodology for implementing the model in a form that can be executed on a computer to calculate the output of a selected algorithm or constitutive equation, and finally studying and visualizing the data generated.

The starting point for a computer simulation involves the selection or development of an idealized or adequate model of the process which is embedded in a mathematical algorithm representing physical laws, state variables, or boundary conditions. When run on a computer, the selected algorithm produces a numerical picture of the system's state as it is conceptualized in the chosen model (Humphreys and Imbert 2010). Computational modeling can provide direction in performing experiments by predicting properties of materials or a series of materials even before their processing or synthesis, or indicate the futility of a synthetic route. Modeling can provide approximate answers to well-characterized materials, while experiments can provide more exact or real answers to materials which are not well characterized. In this sense, computational modeling and simulation can allow for the performance of virtual experiments.

Winsberg (2003) has concluded that, "Successful simulation studies do more than compute numbers. They make use of a variety of techniques to draw inferences from these numbers. Simulations make creative use of calculational techniques that can only be motivated extra-mathematically and extra-theoretically. As such, unlike simple computations that can be carried out on a computer, the results of simulations are not automatically reliable. Much effort and expertise goes into deciding which simulations results are reliable and which are not."

In many simulations, it is not the computing power or time in which the greatest effort is invested but rather in the selection and development of a suitable model which can be verified and validated, often by a lengthy trial-and-error method which alters the input parameters and associated algorithms within sensible ranges describing physical phenomena, especially properties of engineering materials. It is also often found that more time is involved in determining what to simulate than to actually perform the simulations. This is due in part to the fact that for many engineering materials properties, in particular there is no adequate mathematical description to relate them in a quantitative way to microscopic phenomena. This is one of the major limitations of computer simulations in general and computational materials science and engineering in particular. This feature is often described as a length and time-scale problem, but is in reality more of a lack of fundamental knowledge. Bridging the length scale from the atomic or nanoscopic to macroscopic regimes requires a fundamental understanding of the relevant phenomena characteristic of each length scale.

First-principles or ab initio calculations are fundamentally useful since basic data are derived from quantum mechanics (vested in the Schrödinger equation) and therefore require no experimental input on a material. Atomistic-scale computations provide primitive or fundamental information such as energies vested as heat of formation, atom positions, forces, band structures, etc., while engineering concerns are focused on higher-level properties such as ductility, strength, phase stability, resistivity, etc. Engineering properties are usually complex aggregates of the underlying atomic physics, and it is difficult in many instances to relate computational output to anything relevant in an engineering context. Simulations in materials physics are focused on understanding lattice and defect dynamics at the atomic scale, often using *molecular dynamics* (MD) or *Monte Carlo* (MC) *methods*

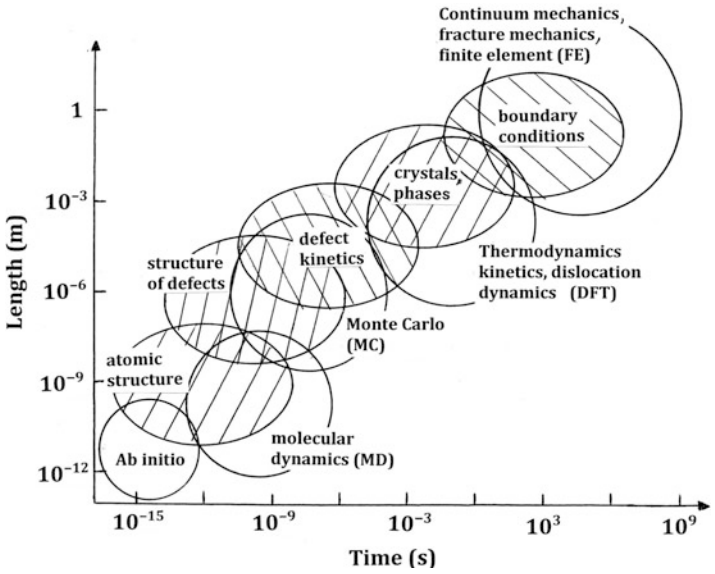


Fig. 2 Key simulation methods for length and time scales applied to materials structures and microstructures

which, as noted, often use physical potentials or force fields derived from solutions of the nonrelativistic Schrödinger equation for some limited number of atoms (Car and Parinello 1985). In contrast, and at the extreme end of the length – time-scale representation shown in Fig. 2, materials-related simulations focused on engineering or large-scale materials problems resort to *finite element methods* (FEM) where the microstructure is homogenized by using averaging, constitutive relationships (Abraham et al. 1998; Bathe 1982; Bulatov et al. 1998) and related methodologies for coupling of length scales (Broughton et al. 1999). Figure 2 illustrates the major simulation methods for different length and time scales superimposed on the hierarchical features of materials microstructures and macrostructures representing materials systems of industrial interest which are highly heterogeneous, and characterized by a variety of defects, interfaces, and related microstructural features (Steinhauser 2008).

In mid-2011, the USA announced *The Materials Genome Initiative*, a new effort to develop an infrastructure to accelerate advanced materials discovery and deployment. One of the major components of The Materials Genome Initiative is the development of toolsets necessary for a new research paradigm where computational analysis and simulation will decrease the reliance on conventional experimentation and testing. In addition data sharing systems and more integrated engineering teams will hopefully allow design and systems engineering as well as manufacturing strategies to be more interactive. This can be illustrated in an integrated materials design continuum model shown schematically in Fig. 3. In this model, the integration of materials computational tools and information

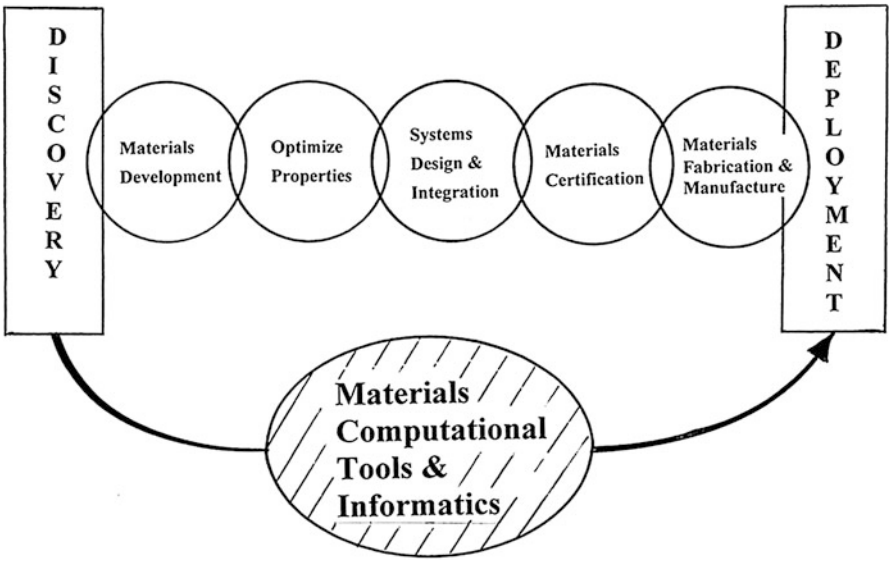


Fig. 3 Materials development model characterizing the materials genome initiative

(or informatics), especially comprehensive materials databases, will drive shorter materials development cycles from the current 10–20 years to roughly one-tenth that time frame. Materials and materials systems models simulating performance, along with appropriate model validation, especially with modular and user-friendly computational software, must be a significant part of developing a materials innovation infrastructure.

Types and Methods of Computer Simulations

Two types of computer simulations are used for three general purposes: prediction, understanding, and exploration. Equation-based simulations are commonly used in the physical or materials sciences where a governing theory can guide construction of mathematical models based on differential equations. These equations are either particle based, involving discrete bodies or differential equation sets governing their interaction, or field based, where a set of equations govern the time evolution of a continuous medium or field. Correspondingly, agent-based simulations are similar to particle-based simulations, but there are no global differential equations governing the motions of individuals being studied in a social or behavioral science context (Epstein 1999). In addition, some simulation models, referred to as multiscale simulation models, couple modeling elements from different description scales. These can involve molecular dynamics (MD), quantum mechanics, or both. Many problems involving classical MD techniques are restricted to small length and time scales as implicit in Fig. 2; an edge length of a few hundred nanometers

simulated for a few nanoseconds, although one timestep in a MD simulation, is typically on the order of femtoseconds ($\sim 10^{-15}$ s). An MD simulation will typically involve several million timesteps and $\sim 10^{10}$ or more particles (Kadau et al. 2004).

Monte Carlo (MC) simulations are another large class of computer simulations or computer algorithms that use randomness to calculate the properties of a mathematical model and where the algorithm randomness is not a feature of the chosen model. Consequently MC simulations are not always simulations of the systems they are being used to study (Grüne-Yanoff and Weirich 2010).

Ab initio methods use quantum mechanics, and especially the Schrödinger equation, to determine the physical properties of molecules. However, the Schrödinger equation is so complex that it can only be solved analytically for a few simple cases involving simple systems and very few atoms. In the time-independent form of the Schrödinger equation:

$$\nabla^2 \psi = [2m(E - V)\psi]/\hbar^2 = 0, \tag{1}$$

$$\psi = \psi(R_1, \ldots . R_N, \ldots r_1 \ldots . r_k, t) \tag{2}$$

where R_i and r_i denote positions of the ith nucleus and the ith electron, respectively. The variable t denotes time. Approximate solutions for the Schrödinger equation in each timestep determine the effective potential energy of the nuclei. Approximate solutions are often obtained by the Hartree–Fock (HF) equation which uses the central field approximation which integrates the electron–electron repulsion, yielding an average effect instead of an explicit energy. A disadvantage of HF calculations is that the effects of one electron on another is not taken into consideration. In contrast, the so-called density functional theory (DFT) takes electron correlation into consideration in order to calculate the ground state energy of many particle systems. Other ab initio methods combine DFT with classical MD in a way that the degrees of freedom of electrons can be treated explicitly in contrast to using classical "effective potentials" between atoms which neglect the electronic movements (Kosloff 1988). DFT uses the electron density rather than the wave function to describe a molecule. With ab initio simulation methods, the only information that must be provided is the number of atoms and their positions within the system. Many codes or software packages are based on DFT, while some implement HF-based models. These include ACES II, QUANTUM ESPRESSO, VASP, and GAUSSIAN. Quantum mechanical calculations often provide a connection between the atomistic and microscopic scales (Fig. 2) by using effective interaction potentials including variations of MD and ML methods such as the discrete element method (DEM) (Cundall and Strack 1979) or dissipative particle dynamics (DPD) (Hoogerbrugge and Koelman 1992). With these methods individual particles are represented by complete clusters of atoms or molecules and are treated as classical particles which can consist of millions of atoms. Although this approach does not provide an atomistic description of molecular motion, it does represent the correct hydrodynamic behavior on long length and time scales.

Both MC and MD simulations can generate representative ensembles; MC methods are generally simpler than MD methods and do not require molecular force calculations. However, for dynamic events and the generation of nonequilibrium ensembles, MD is more appropriate. MD simulations in their simplest form involve the step-by-step numerical solution of classical Newtonian equations of motion:

$$F_i = m_i \frac{d^2 r_i}{dt^2} = -\nabla\phi(r_i - r_N), \tag{3}$$

corresponding to N particles of mass m_i having position vectors r_i, with interaction potentials, ϕ, and force fields – $\nabla\phi$. Starting from the original Lennard–Jones potential between two particles i and j with distance $r = (r_i - r_j)$ which has the general form

$$\phi_{a,b}(r) = \alpha\varepsilon\left[\left(\frac{\sigma_o}{r}\right)^a + \left(\frac{\sigma_o}{r}\right)^b\right], \tag{4}$$

where

$$\alpha = \frac{1}{a-b}\left(\frac{a^a}{b^b}\right)^{\frac{1}{a-b}},\ \phi_{min} = \varepsilon \text{ and } \phi(\sigma) = 0 \tag{5}$$

the most commonly used Lennard–Jones (LJ) potential for the interaction between two particles with a distance $r = (r_i - r_j)$ is expressed by

$$\phi_{LJ}(r) = 4\varepsilon\left[\left(\frac{\sigma_o}{r}\right)^{12} + \left(\frac{\sigma_o}{r}\right)^{6}\right]. \tag{6}$$

Here ε determines the energy scale, while σ_o determines the length scale. Considering Eq. 6, the potential function for N interacting particles becomes

$$\phi(r_1 \ldots r_N) = \sum_{i=1}^{N}\sum_{j=i+1}^{N}\phi_{LJ}(r) = 4\varepsilon\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(\frac{\sigma_o}{r}\right)^6 \times \left(\left(\frac{\sigma_o}{r}\right)^6 - 1\right). \tag{7}$$

The corresponding force, F_i, exerted on particle i by particle j is then given by the gradient with respect to r_i as

$$F_i = -\nabla_{ri}\phi_{LJ}(r_i \ldots r_N) = -24 \times \varepsilon\sum_{j=1}^{N}\sum_{j\neq 1}^{N}\frac{1}{r^2}\left(\frac{\sigma_o}{r}\right)^6 \times \left(1 - 2\left(\frac{\sigma_o}{r}\right)^6\right)r_{ij}. \tag{8}$$

where $r_{ij} = (r_i - r_j)$ is the direction vector between particles i and j at positions r_i and r_j, and $r = (r_i - r_j)$. The force, F_i, or particle i is then the sum over all forces, $F_{ij} = -\nabla_{ri}\phi$ between particle i and all other particles j:

$$F_i = \sum_{i=1, j\neq 1}^{N} F_{ij} \tag{9}$$

Equations 6, 7, 8, and 9 represent the most crucial part of an MD simulation which has to be optimized in terms of the force calculation which determines the interacting particle pairs. Correspondingly, the analytical form of the potential which is derived from theory is then consistently adjusted to experimental findings (Maitland et al. 1981).

In the mesoscopic scale, methods based on continuum theory are used. These involve mesh-based methods such as finite element methods (FEM). Smooth particle hydrodynamics (SPH) represents a modern continuum method based on the conservation equations of continuum theory which avoids mesh distortion problems inherent in FEM approaches (Liu and Liu 2003).

Continuum models which describe elastic and viscoelastic behavior of solids and fluids dominate macroscale methods of simulation, particularly the discretization of the continuum into discrete elements using FEM where elements are connected together by a topological map or mesh. Here finite element interpolation functions are built upon the mesh which insures compatibility of the interpolation. Hydrocode or the so-called wave propagation codes are also utilized which decouple the stress tensor in a deviatoric and hydrostatic component. These methods are based on a solution of the continuum conservation equations of mass, momentum, and energy as well as explicit formulations of equations of state using the so-called constitutive equations or relationships. ANSYS AUTODYN software, LS-DYNA, and ABACUS software are often used in simulations involving these issues as applied to shock and impact simulation illustrated in chapter "▶ Ballistic and Hypervelocity Impact and Penetration." These software methods encompass a number of different numerical approaches for impact-related problems in particular, which include Lagrange, Euler, arbitrary Lagrange–Euler (ALE), Shells, and SPH numerical processors (Cook et al. 2001). These codes also contain conservation equations, constitutive equations, and failure modes which deal with fracture, spallation, and related dynamic phenomena such as dynamic recrystallization (DRX). Lagrangian codes are characterized by a referential where the computational grid deforms with the material, while in an Eulerian referential it is fixed in space. While Lagrangian codes are conceptually more straightforward, the Eulerian approach is preferential for problems involving large deformations.

It might be noted that in FEM, forces, stresses, and strains are related by writing equations in matrix form, while in finite difference (FD) methods, the derivatives or gradients are replaced by simple differences. For example, in FD a slope of a tangent derivative is replaced by a simple formula of slope of a straight line.

The plastic response in uniaxial loading can be expressed in a simple constitutive equation of the form:

$$d\sigma = \left(\frac{\partial \sigma}{\partial \varepsilon}\right)_{\dot{\epsilon}, T} \partial \varepsilon + \left(\frac{\partial \sigma}{\partial \dot{\varepsilon}}\right)_{\epsilon, T} \partial \dot{\varepsilon} + \left(\frac{\partial \sigma}{\partial T}\right)_{\varepsilon, \dot{\varepsilon}} dT \tag{10}$$

which connects incremental changes in true axial stress, σ, to changes in true plastic axial strain, ε; true plastic axial strain rate, $\dot{\varepsilon}$; and absolute temperature, T. The partial derivatives shown in Eq. 10 represent materials parameters and depend on the microstructure, which in turn is influenced by deformation history of the

material. In this sense, the most common representation of the plastic response of a material or the stress–strain curve of a polycrystalline metal in particular is given by the Ludwik–Hollomon equations:

$$\sigma - K\varepsilon^{n} \tag{11}$$

and

$$\sigma = \sigma_o + K\varepsilon^{n}, \tag{11a}$$

where σ_o and K are constants and the exponent n is called the work-hardening coefficient. More general forms of Eq. 11a incorporating strain rate and thermal effects implicit in Eq. 10 are expressed in physically based constitutive equations such as the Johnson–Cook equation (Johnson and Cook 1983):

$$\sigma = (\sigma_o + K\varepsilon^{n})\left(1 + C\ln\left(\frac{\dot{\varepsilon}}{\dot{\varepsilon}_o}\right)\right)\left[1 - \left(\frac{T - T_r}{T_m - T_r}\right)^{m}\right], \tag{12}$$

where K, C, n, and m are material parameters, T_r is a reference temperature, T_m is the melting point, and $\dot{\varepsilon}_o$ is the reference strain rate. The three groups of terms in Eq. 12 represent the work-hardening (Eq. 11a), strain rate, and thermal effects, respectively. There are other constitutive equations which incorporate microstructural issues such as grain size and dislocation effects. Zerilli and Armstrong (1992) developed two microstructurally based constitutive equations based on the framework of thermally activated dislocation motion in the context of temperature and strain-rate response of FCC and BCC metals:

$$\left.\begin{aligned} \textbf{FCC}: \sigma &= \sigma_G + C_2\varepsilon^{0.5}\exp(-C_3T + C_4T\ln\dot{\varepsilon}) + KD^{0.5} \\ \textbf{BCC}: \sigma &= \sigma_G + C_1\exp(-C_3T + C_4T\ln\dot{\varepsilon}) + C_5\varepsilon^{n} + KD^{0.5} \end{aligned}\right\} \tag{13}$$

where σ_G and C_i (i = 1 to 5) are material constants, T is the temperature, and D is the grain size. The difference between the FCC and BCC equations (Eq. 13) is that the plastic strain is uncoupled from the strain rate and temperature for BCC metals.

Table 1 provides a brief overview of some of the more prominent and different simulation techniques and software packages used on various length scales in materials and chemical sciences, along with representative applications. This overview represents a spatial rather than a physical classification. These software systems and hydrocodes can be 1D, 2D, 3D, or in combination. Many are 2D and 3D, including the many FE and FD methods.

One of the more significant modeling methods has involved density functional theory (DFT) applications in predicting properties of crystalline materials such as energies of formation, lattice parameters, band structures, etc. The so-called high-throughput (HT) DFT calculations have generated large databases of DFT-predicted materials properties, including the Open Quantum Materials Database (DQMD) (available at http://oqmd.org), which currently contains predicted crystallographic parameters and formation energies for over 200,000 structures, and the Inorganic Crystal Structure Database (ICSD) (Saal et al. 2013).

Table 1 Simulation methods and software examples associated with length scale applications[a]

Scale (m)	Simulation methods/software	Application examples
Electronic/atomistic		
$\sim 10^{-12}$–10^{-9}	Self-consistent Hartree–Fock (SC-HF) and self-consistent DFT/CRYSTAL, GAUSSIAN, Q-CHEM, VASP (Phillips 2003; Steinhauser 2008), ab initio MD/MOLPRO Quantum MC (Nightingale and Umrigar 1999)	Crystal ground state, molecular structures and geometry, electronic properties, chemical reactions, energy of formation, structural energetics
Atomistic/microscopic		
$\sim 10^{-9}$–10^{-6}	MD; MC using classical force fields; Hybrid MD/MC Embedded atom method (EAM) (Daw 1988; Steinhauser 2008)	Equations of state, rheology, transport properties, phase equilibrium and phase stability
Microscopic/mesoscopic		
$\sim 10^{-8}$–10^{-1} ↓ $\sim 10^{-6}$–10^{0}	MD and MC using effective force fields (Steinhauser 2008). Dissipative particle dynamics (Hoogerbrugge and (Hoogerbrugge and Koelman 1992). Finite element methods (FEM) including microstructural features (Phillips 2001)/ABACUS. SPH, dislocation dynamics, discrete element method (DEM) (Cundall and Strack 1979)	Granular matter, fracture mechanics, grain growth and grain boundaries, polycrystal elasticity and plasticity, phase transformations, diffusion, dislocations, motion of interfaces
Mesoscopic/macroscopic		
$\sim 10^{-3}$–10^{2} ↓ $\sim 10^{-6}$–10°	Hydrodynamics, computational fluid dynamics; FEM/ANSYS AUTODYN, DYNA FDM/WONOY, HEMP; Percolation and cluster models (Benson 1992)	Macroscopic flow, elasticity and plasticity; fracture mechanics; fatigue and wear; impact and penetration; shock wave phenomena

[a]Adapted from Steinhauser and Hiermaier (2009). See text for method abbreviations

It is also important to briefly discuss computational methods applied to materials thermodynamics and especially thermodynamic phase stability applications of first-principles calculations, including DFT. These relate to the advent of the calculation of phase diagrams (CALPHAD) method (Wang et al. 2013), especially modern versions of PANDAT software (Cao et al. 2009) for multicomponent phase diagram calculation and materials property simulation. Cao et al. (2009) have recently described the CALPHAD approach to require the information of the Gibbs free energies of all phases as functions of temperature and chemical composition, where the general form of the Gibbs free energy for a phase α in the CALPHAD approach is expressed as

$$G^{\alpha} = {}^{o}G^{\alpha} + \Delta G^{\alpha}, \tag{14}$$

where G^{α} is the Gibbs free energy of the multicomponent α phase, ${}^{o}G^{\alpha}$ represents the mechanical mixing of individual species, and ΔG^{α} represents interactions among the species or endmembers. Following this approach,

$$ {}^{o}G^{\alpha} = \Sigma\, x_i {}^{o}G_i{}^{\alpha} \tag{15} $$

and

$$ \Delta G^{\alpha} = \Delta G^{\alpha}{}_{conf} + \sum_{i} \sum_{j>i} x_i x_j \sum\nolimits_{n=0}^{n} L_{ij}^{\alpha} \left(x_i - x_j\right)^n + \sum_{i} \sum_{j>i} \sum_{k>j} x_i x_j x_k L_{ijk}^{\alpha}; \tag{16} $$

where x_i is the mole fraction of species i, ${}^{o}G_i{}^{\alpha}$ is the Gibbs free energy of species i in the structure of phase α, $\Delta G^{\alpha}{}_{conf}$ is the configurational contribution, and ${}^{n}L^{\alpha}{}_{ij}$ and ${}^{n}L^{\alpha}{}_{ijk}$ are binary and ternary interaction parameters which can be temperature dependent. $\Delta G^{\alpha}{}_{conf}$ is usually assumed to be from the ideal atomic configurational entropy of mixing given by

$$ -T\Delta S_{conf} = RT \sum x_i \ln x_i, \tag{17} $$

where R is the gas constant, T is temperature, and ΔS_{conf} represents the ideal atomic configurational entropy. In modeling multicomponent atomic mobility, molar volume, and elastic coefficients using CALPHAD, the configurational contribution represented in Eq. 17 can be neglected (Wang et al. 2013).

Computational methods such as MD and DFT have also been applied to investigations of theoretical and practical aspects of solidification phenomena and simulation of solidification across the length scales where FEM is also used to investigate the die-filling and solidification processes involved with a wide range of components spanning from automotive to aero engine turbine blades (Dantzig and Rappaz 2009). At the atomic level, MD and DFT can investigate the solid/liquid interface and its effect on microstructural morphology, while FEM and related simulations attempt to bridge the gap between solidification pattern formation at the scale of dendrite and process models involved with thermal transport and mass transport at the mesoscale and above.

It should also be pointed out that practical simulations, or engineering simulations, involve a wide range of computer-aided design (CAD) software such as Materialise and SolidWorks described and illustrated in chapter "▶ Rapid Prototyping Technologies: Solid Freeform Fabrication" in connection with additive manufacturing involving the array of 3D printing technologies. SolidWorks Simulation is an FEM to calculate component displacements, stresses, and strains under internal and external loads. This software is tightly integrated with SolidWorks 3D CAD so that geometry to be analyzed is discretized using tetrahedral (3D), triangular (2D), and beam elements, and solved by either a direct sparse or iterative solver involving linear or nonlinear stress analysis.

While this section has presented an overview of computer simulation methods, there has been no attempt to provide any tutorial-like descriptions of either these methods or their arrays of applications in materials science and engineering.

For more detailed descriptions and applications, the reader is referred to the recent introductory overview by LeSar (2014) as well as the description and application examples provided in Phillips (2003), and Steinhauser (2008). The following section of this chapter will illustrate a few examples of implementing simulation methods as well as issues related to verification and validation of simulations and simulation methods.

Implementation, Verification, and Validation of Simulations

It is often said that *simulations without validation are nothing more than animation*. While validation on the one hand involves the process of determining whether the selected model represents the real situation, verification, on the other hand, involves the process of determining whether the simulation approximates true solutions to the differential equations representing the original model. Validation involves comparing a model output with experimental or observable data. Roy (2005) has posed that, "verification deals with mathematics and addresses the correctness of the numerical solution to a given model. Validation, on the other hand, deals with physics and addresses the appropriateness of the model in reproducing experimental data. Verification can be thought of as solving the chosen equations correctly, while validation is choosing the correct equations in the first place." In practice, the successful implementation of a model or simulation method is determined by trial-and-error adjustment between the model and the method of calculation. Consequently, it is often difficult to separate verification and validation in achieving a successful simulation. Indeed, a simulation which accurately depicts complex phenomena invariably contains a wealth of information related to that phenomena. Winsberg (2010) has recently argued that when there is relevant background knowledge in place, a simulation can actually provide more reliable knowledge of a system than an experiment. In effect, simulations can run a large number of experiments involving many variables, but there must be some basis for validation; otherwise, the experiments represented by the simulations will be largely fictitious and unrepresentative of the real world or real materials, where materials are the focus.

In chapter "► Ballistic and Hypervelocity Impact and Penetration," Fig. 14, a 3D computer simulation sequence was presented for a 3.18 mm diameter stainless steel (SS) projectile impacting a thick (semi-infinite) Ni target at a velocity of 3.4 km/s. It might be instructive to illustrate how this simulation or simulation sequence was implemented using a 3D AUTODYN (version 5.0) software package and subsequently validated (and verified). The starting point involved the actual impact experiments utilizing 3.18 mm diameter (spherical) 440 stainless steel projectiles launched at different velocities from either a powder propellant gun or gas gun at normal incidence against 2.5 cm thick stationary Ni targets. These impacted targets were cut in half and observed in the scanning electron microscope (SEM) to examine the fragmentation of the projectile and deposition within the excavated crater, at different impact velocities (Hernandez, et al. 2006).

The AUTODYN computer simulation involved first selecting processors corresponding to the projectiles and the Ni target. The smoothed particle hydrodynamic (SPH) processor with a simple principal stress failure model used to model the projectile was coupled with a Lagrangian target model where the grid distorts with the target material. An erosion algorithm was also implemented to alleviate grid tangling and element degeneracy. This algorithm works by removing Lagrangian cells that have reached a user-prescribed strain (Hayhurst et al. 1995). SPH was used to model the projectile because it was assumed that it would undergo very large deformation without the need for additional algorithms such as an erosion algorithm.

A polynomial equation of state was chosen for the Ni target, while the shock equation of state was used for the 440 SS projectile:

$$U_s = C_o + SU_p, \tag{18}$$

where U_s is the shock velocity, C_o is the sound velocity, S is an empirical parameter related to the Grüneisen parameter, and U_p is the particle velocity. The Johnson–Cook strength model (or constitutive equation (Eq. 12)) was applied to the SS projectile in the form:

$$\sigma = \left[A + B\varepsilon_p{}^n\right]\left[1 + C \ln \dot{\varepsilon}_p{}^*\right]\left[1 - T_H{}^m\right], \tag{19}$$

where ε_p is the effective plastic strain, $\dot{\varepsilon}_p{}^*$ is the normalized effective plastic strain rate (Eq. 12), and T_H is the homologous temperature defined as

$$T_H = (T - T_{room})/(T_{melt} - T_{room}), \tag{20}$$

and A, B, C, n, and m are material constants given in the AUTODYN software material library.

The Ni targets were modeled using the Steinberg–Guinan strength model where the constitutive relationships for the shear modulus, G, and the yield stress, σ, at high strain rates, are given by (Steinberg et al. 1980):

$$G = G_o\left[1 + \left(G'_p/G_o\right)P/\eta^{1/3} + \left(G'_T/G_o\right)(T - 300)\right] \tag{21}$$

and

$$\sigma = \sigma_o\left[1 + \left(\sigma'_p/\sigma_o\right)P/\eta^{1/3} + \left(G'_t/G_o\right)(T - 300)(1 + \beta\varepsilon)^n\right], \tag{22}$$

subject to $\sigma_o\ (1 + \beta\varepsilon)^n \leq \sigma_{max}$, where ε is the effective plastic strain, T is the temperature (K), η is the compressibility equal to V_o/V, and the primed parameters with subscripts p and T are derivatives of that parameter with respect to pressure (P) and temperature (T) at the reference state ($T = 300$ K, $P = 0, \varepsilon = 0$). The subscript zero also refers to values of G and σ at the reference state. Values for these parameters are also given in the AUTODYN software materials library as input parameters. These are listed for the 440 SS projectile and Ni targets shown in Table 2.

Table 2 Material input parameters for 3D impact crater simulation[a]

Parameter	Stainless steel Projectile	Nickel target
Equation of state	Shock	Polynomial
Strength model	Johnson–Cook	Steinberg–Guinan
Steinberg–Guinan		
Reference density (g/cm^3)	7.86	8.9
Reference temperature (K)	300	300
Specific heat (Terg/gK)	4.77E-06	4.01E-06
Shear modulus (Mbar)	8.18E-01	6.50E-01
Yield stress (Mbar)	1.00E-01	1.70E-03
Maximum yield stress (Mbar)	3.00E-02	
Hardening constant (Mbar)	20	4.60E + 01
Hardening exponent	2.50E-01	5.30E-01
Strain rate constant	1.20E-02	
Thermal softening exponent	7.50E-01	
Melting temperature (K)	2,370	2,330
Failure model	Principal stress	None
Erosion (instantaneous)	None	Inst.geometrical
Erosion strain		1.5

[a]After Hernandez et al. (2006)

Figure 4a illustrates the 440 SS projectile and Ni target 3D-half-section model for the computer-simulated impact and their corresponding processors. Figure 4b shows a 3D-half-section SEM view for a 440 SS projectile impacting a Ni target at 3.4 km/s observed in the SEM showing the SS fragments in the crater wall. Correspondingly, Fig. 4c–f shows SEM 3D-half-section views for SS projectiles impacting Ni targets at velocities indicated above, where Fig. 4f corresponds to Fig. 4b. In Fig. 4g–j, the computer-simulated 3D-half-section views corresponding to Fig. 4c–f are reproduced. Simply comparing the real SEM 3D-half-section views in Fig. 4c–f and the computer-simulated views in Fig. 4g–j provides a high level of verification and simulation validation. By considering the crater geometry or crater depth/crater diameter ratio (P/D_c), simple numerical comparisons can also be made as shown in Table 3 where the largest variance of 7 % is observed for the 3.4 km/s impact corresponding to Fig. 4f and j, or 93 % validation based on this simple ratio.

Although Fig. 4 illustrates a relatively simple example of computer simulation validation related to materials, it provides a broader perspective for simulations which can relate to time or events within some time frame, as shown in ▶ Fig. 14 of chapter "Ballistic and Hypervelocity Impact and Penetration" which illustrates the formation of the crater in Fig. 4f–j. This feature of computer simulation is particularly unique and especially useful in the context of dynamic events such as impact since in reality there are only two end points: the projectile striking the target surface at time, t equal zero, and the fully formed crater at some steady-state time. Consequently, validation of a computer simulation for a particular materials system

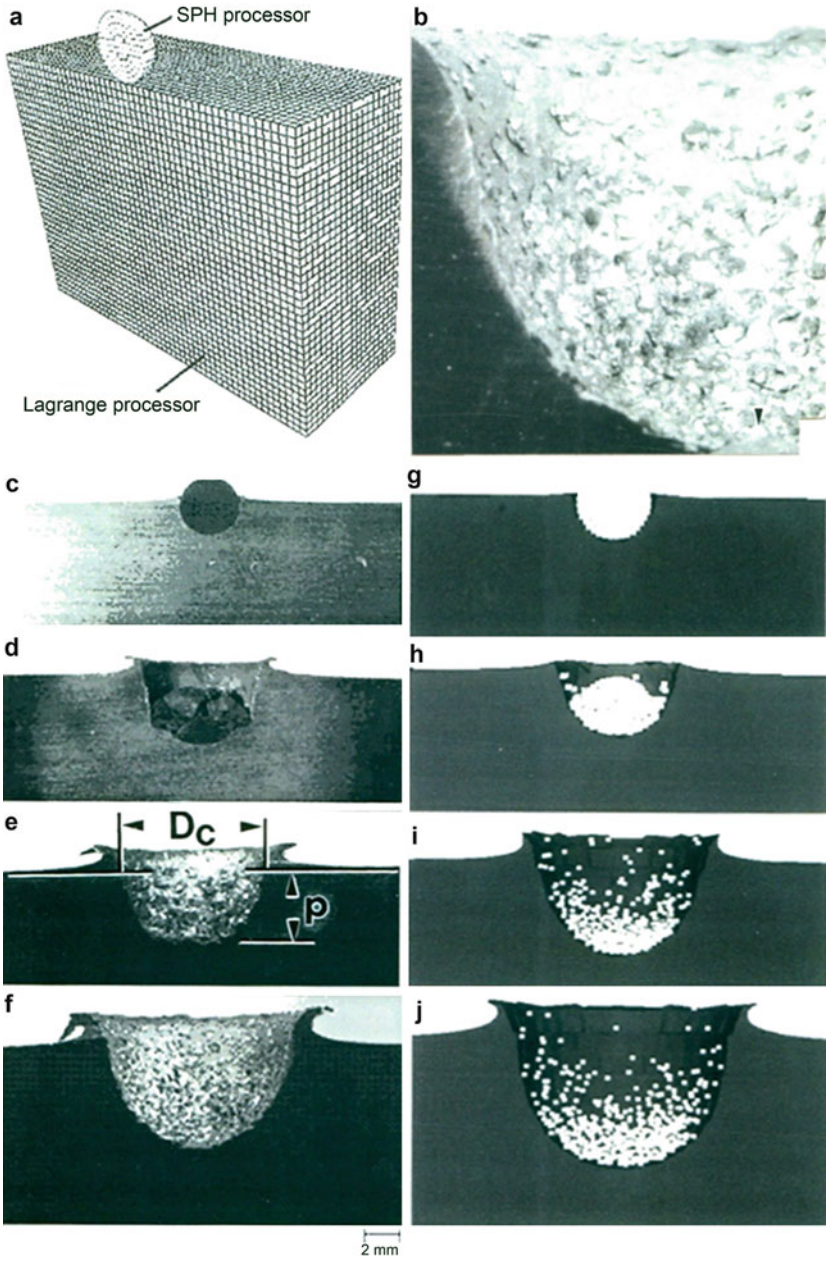


Fig. 4 (**a**) Projectile/target 3D-half-section model for computer-simulated impact. (**b**) SEM view of 3D-half-section of Ni crater with 440 SS projectile fragments after impact at 3.4 km/s. (**c–f**) show experimental, SEM 3D-half-section views at impact velocities of 0.52, 1.1, 2.2, and 3.4 km/s, respectively. (**g–j**) show computer-simulated 3D-half-section views at impact velocities of 0.52, 1.1, 2.2, and 3.4 km/s respectively (Adapted from Hernandez et al. (2006))

Table 3 Experimental and simulated impact crater data based on Fig. 4c–j for 440 SS projectiles impacting Ni targets[a]

Impact velocity, U_o (km/s)	Experimental (p/D_c)	Simulated (p/D_c)
0.52	0.52	0.52
1.10	0.59	0.60
2.20	0.61	0.62
3.4	0.57	0.61

[a]After Hernandez et al. (2006)

can allow for the detailed event sequence to be examined in some time interval or time sequence. These features are illustrated for related impact events in ▶ Figs. 14, ▶ 24, ▶ 26d, and ▶ 28 of chapter "Ballistic and Hypervelocity Impact and Penetration". ▶ Figures 22, ▶ 26, ▶ 28a–c, and ▶ 29 of chapter "Ballistic and Hypervelocity Impact and Penetration have also shown examples of 2D computer simulation validation for related impact events in materials systems. It must be cautioned that valid simulations for one materials system is not necessarily applicable or able to be extrapolated to another different materials system, although, as illustrated in Fig. 4, it may be possible to validate only one or a few experimental events to allow the simulation to be applied in other circumstances involving initial impact or related physical conditions.

References

Abraham F, Broughton J, Bernstein N, Kaxiras E (1998) Spanning the length scales in dynamic simulation. Comput Phys 12:538–556

Bathe K (1982) Finite element procedures in engineering analysis. Prentice Hall, Cambridge, UK

Benson DJ (1992) Computational methods in Lagrangian and Eulerial hydrocodes. Comp Methods in Appl Mech & Engr 99(2–3):235–394

Broughton JQ, Abrahams FF, Bernstein N, Kaxiras E (1999) Concurrent coupling of length scales; methodology and applications. Phys Rev B60:2391–2403

Bulatov VU, Abraham F, Kubin L, Devrince B, Yip S (1998) Connecting atomistic and mesoscale simulations of crystal plasticity. Nature 391:669–672

Car R, Parinello M (1985) Unified approach for molecular dynamics and density-functional theory. Phys Rev Lett 55:2471–2474

Cao W, Chen S-L, Zhang F, Wu K, Yang Y, Chang YA, Schmid-Fetzer R, Oates WA (2009) PANDAT Software with PanEngine, PanOptimizer and PanPrecipitation for multi-component phase diagram calculation and materials property simulation. CALPHAD Comp Coupling Phase Diagrams Thermochem 33:328–342

Cook RD, Malkus DS, Plesha ME, Witt RJ (2001) Concepts and applications of finite element analysis, 4th edn. Lavoisier SAS, Paris

Cundall P, Strack O (1979) A discrete numerical model for granular assemblies. Geotechnique 29:47–65

Dantzig JA, Rappaz M (2009) Solidification. CRC Press/Taylor and Francis Group, LLC, Boca Raton

Daw MS (1988) Model of metallic cohesion: the embedded atom method. Phys Rev B 39:7441–7452
Epstein J (1999) Agent-based computational models and generative social science. Complexity 4(5):41–57
Grüne-Yanoff T, Weirich P (2010) Philosophy of simulation. Simul Gaming Interdiscipl J 41(1):1–31
Hayhurst CJ, Ranson HJ, Gardner DJ, Birnbaum NK (1995) Modeling of microparticle hypervelocity oblique impacts on thick targets. Int J Impact Eng 17:375–386
Hernandez VS, Murr LE, Anchondo IA (2006) Experimental observations and computer simulations for metallic projectile fragmentation and impact crater development in thick metal targets. Int J Impact Eng 32:1981–1999
Hoogerbrugge P, Koelman J (1992) Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. Europhys Lett 19:155–160
Humphreys P, Imbert C (eds) (2010) Models, simulations and representations. Routledge Publishers, London
Johnson GR, Cook WH (1983) A constitutive model and data for metals subjected to large strains, high strain rates and temperatures. In: Proceedings of the 7th international symposium. O Ballistics, The Hague
Kadau K, Germann T, Lomdahl P (2004) Large-scale molecular dynamics simulation of 19 billion particles. J Modern Phys C 15:193–201
Kosloff R (1988) Time-dependent quantum-mechanical methods for molecular dynamics. J Chem Phys 92:2087–2100
Le Sar R (2014) Introduction to computational materials science. Cambridge University Press, Cambridge, UK
Liu GR, Liu MB (2003) Smoothed particle hydrodynamics. A meshfree particle method. Scientific, Singapore
Maitland G, Rigby M, Smith E, Wakeham W (1981) Intermolecular forces – their origins and determination. Clarendon Press, Oxford
Nightingale M, Umrigar C (eds) (1999) Quantum Monte Carlo methods in physics and chemistry. Springer, New York
Phillips R (2003) Crystals, defects and microstructures – modeling across scales. Cambridge University Press, Cambridge, UK
Roy S (2005) Recent advances in numerical methods for fluid dynamics and heat transfer. J Fluid Eng 127(4):629–630
Saal JA, Kirklin S, Aykol M, Meredig B, Wolverton C (2013) Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). JOM 65(1):1501–1509
Steinberg DJ, Cochran SG, Guinan MW (1980) A constitutive model for metals applicable at high strain rates. J Appl Phys 51(3):1498–1502
Steinhauser MO (2008) Computational multiscale modeling of solids and fluids – theory and applications. Springer, Heidelberg
Steinhauser MO, Hiermaier S (2009) A review of computational methods in materials sciences: examples from shock-wave and polymer physics. Int J Mol Sci 10:5135–5216
Wang Y, Shang S, Chen L-Q, Liu Z-K (2013) Density functional theory-based database development and CALPHAD automation. JOM 65(1):1533–1539
Winsberg E (2003) Simulated experiments: methodology for a virtual world. Phil Sci 70:105–125
Winsberg E (2010) Science in the age of computer simulation. The University of Chicago Press, Chicago
Zerilli FJ, Armstrong RW (1992) The effect of dislocation drag on the stress–strain behavior of fcc metals. Acta Met Mater 40:1803–1809

Applications and Examples of Multiscale Computer Simulations in Materials Science and Engineering

Contents

Abstract

As the culmination of this handbook, this chapter represents more than 12 orders of magnitude in length scales (in meters) and more than 25 orders of magnitude in timescales (in seconds) representing materials structures, properties, processing and performance issues, and phenomena as these are related to computer simulations and especially example applications. These illustrate atomistic simulations and simulations applied to large engineering structures such as the Francis hydraulic turbine runner blade castings for the Three Gorges Dam (China), weighing over 400 t. These examples also illustrate the concepts of verification and validation essential in successfully applying computational materials science and engineering to the widest spectrum of materials and materials systems.

Introduction

As illustrated in chapter "▶ Rapid Prototyping Technologies: Solid Freeform Fabrication," computer simulation as it relates to computer-aided design (CAD) software allows complex materials system development and fabrication to be visualized in the same way architectural designs illustrate and visualize a building or other structure. In principle, it should be possible to design a new material or a materials system in the same way by optimizing composition and processing steps

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7_61

to achieve desired structures (microstructures) and associated properties required for some intended application or performance features. Since materials properties are ideally determined by the constituent atoms and fundamental laws of physics, atomic-scale engineering of materials properties based on ab initio calculations, where basic data are derived from quantum mechanics, holds considerable promise because they require no experimental material input. In addition, predicting properties of a series of materials or materials systems prior to actual synthesis or processing by computer modeling and synthesis can guide experiments in promising directions. Ideally, materials or materials systems visualized in a computer simulation can be perfectly characterized in terms of structure, composition, properties, etc., making the computational modeling and simulation often less ambiguous than actual experiments. Once validated, computer simulations can, as illustrated in ▶ Fig. 4 of chapter "Computer Simulation in Materials Science and Engineering," be repeated for a large number of materials systems, allowing virtual experiments to be performed to evaluate different materials and their compositional modifications which can allow theories to be developed based on the results.

There are still many important materials properties which are difficult to model, a fact that imposes serious limitations on the ability to apply computer simulations. In addition, ab initio or atomistic scale computations provide limited, fundamental information such as atom positions, forces, energies, etc., while in materials engineering applications higher level properties such as mechanical behavior, corrosion resistance, phase stability, etc., are of more interest. But since many engineering properties do not have an adequate mathematical description relating them to materials microscopic phenomena in a quantitative way, this poses limitations on computational materials simulations especially in the context of length and timescale phenomena. Indeed, bridging the length scale from microscopic to macroscopic as implicit in ▶ Fig. 2 and Table 1 of chapter "Computer Simulation in Material Science and Engineering" requires an understanding of relevant phenomena on each scale. This is especially important in implementing integrated computational materials engineering simulations involving validated computational approaches that span vast differences in length and timescales and allow efficient transfer and incorporation of simulation-based materials science and engineering in industry.

This chapter illustrates a few examples of applications of ab initio and related fundamental computer simulations as well as broader, practical materials engineering applications of computer simulations spanning the length and timescales represented in ▶ Fig. 2 and Table 1 of chapter "Computer Simulation in Materials Science and Engineering."

Atomistic Simulations

As illustrated in chapter "▶ Examples of Materials Science and Engineering in Antiquity," Fig. 4, Lair et al. (2006) have shown using ab initio simulations that carbon nanotube caps, or end caps, play an important role in modeling since the

specific placement of hexagons and pentagons in the end cap will create an array of single-wall carbon nanotubes (SWCNTs) with predictable geometries related to specific energies or heats of formation. However, more recent ab initio simulations involving double-wall carbon nanotube (DWCNT) formation and geometries related to electronic energies, energy trends, and DWCNT stability by Lair et al. (2008) performed at a variety of computational levels have revealed a fundamental understanding and clarification of nucleation and favored (energetically favorable) growth paths. A few examples of this work not only illustrate these atomistic phenomena but also illustrate that in computations involving atomic energies or heats of fusion, validation is not a significant issue since energy variations or systematic comparisons of energy calculations will allow for decisions regarding process stability or favorability based on energy minimization, a measure of verification.

The majority of carbon nanotube geometries in this example were optimized using semiempirical Austin Model l (AM1) as Parametric Method Number 3 (PM3) calculation methods (Lair et al. 2008). Electronic energies were determined with the Hartree-Fock (HF) level of theory using STO-3G, 6-31G, or 6-311G basis sets. STO-3G uses the minimal number of functions to describe each atom in a molecule which for nanotube systems means that hydrogen is described by one function (for the 1 s orbital) while carbon is described by five functions: 1 s, 2 s, $2p_x$, $2p_y$, and $2p_z$. 6-31G or 6-311G (which uses three functions for each orbital) includes more functions to describe different sized orbitals. Using GAUSSIAN 03 W software, energies were given in atomic units (a.u.), also known as Hartrees: 1 a.u. = 1 Hartree = 627.51 kcal/mol. These basis sets and simulation models were used to optimize computational time and accuracy required for a large number of calculations and corresponding simulations. Density functional theory (DFT) calculations were also performed on some systems in order to compare accuracies and provide some measure of simulation verification in the context of energetic trends and especially double-wall carbon nanotube (DWCNT) growth characteristics, including nucleation or cap structure and energetics, growth stage energetics, and minimum energy or stability configurations of different carbon nanotube types: armchair (A) versus zigzag (Z) with reference to the graphene lattice conventions illustrated in chapter "▶ Examples of Materials Science and Engineering in Antiquity," Fig. 5.

Figure 1a illustrates DWCNT geometries represented by an inner (initial) CNT cap (N1) nucleating a carbon nanotube having diameter, d, given by $d = a\,(n^2 + nm + m^2)^{½}$, where a = 2.46 Å, and an outer CNT cap (N2) nucleating the outer carbon nanotube which is separated from the initial, inner CNT by an interwall spacing, $(IWS) = (d2-d1)/2-r_o-r_i$, where r_o is the outer CNT radius and r_i is the initial (inner) CNT radius. Figure 1b shows a simulated DWCNT zigzag (inner tube)/zigzag (outer tube) section: Z/Z designated (6, 0) @ (13, 0) (▶ Fig. 5 of chapter "Examples of Materials Science and Engineering in Antiquity") with no end caps: three rows of outer hexagons (six-member rings) and two rows of inner hexagons (six-member rings) terminated in hydrogen simulated using PM3 for geometry optimization and the HF STO-3G model and basis set for single-point energy optimization. Figure 1c shows this zigzag (Z) configuration schematically, while

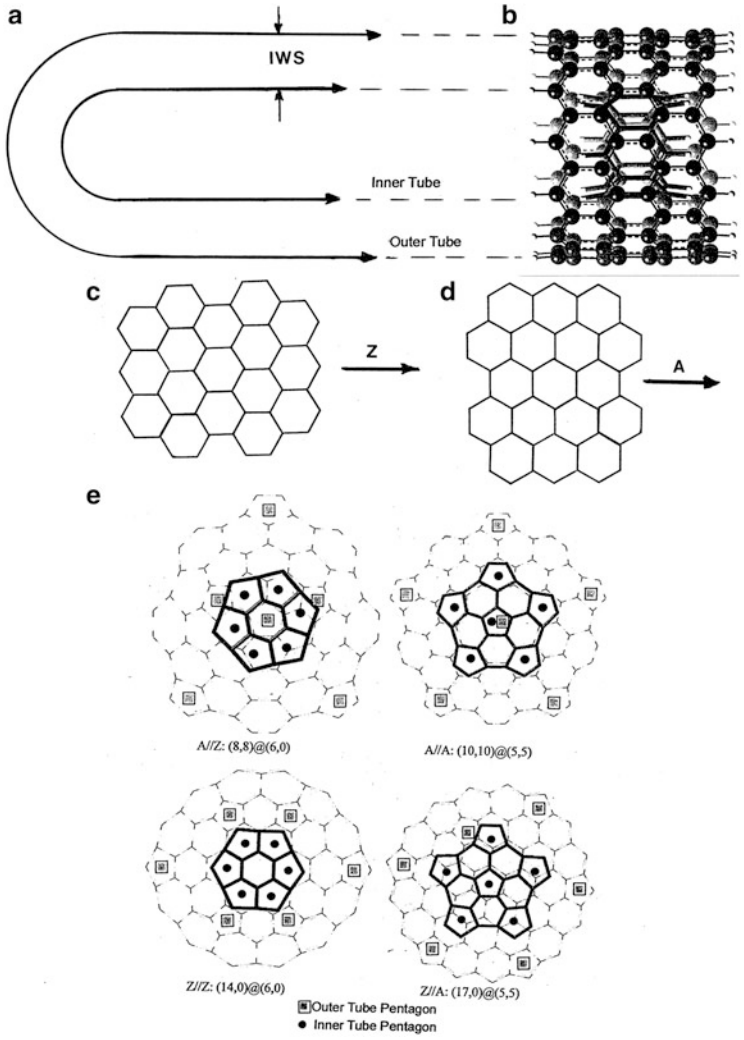


Fig. 1 (**a**) Schematic representation of tube diameter outlines for DWCNT structure. (**b**) Example of Z/Z: (6, o) @ (13, o) DWCNT simulation with no end caps; terminated by hydrogen (H) atoms. (**c**) Zigzag (*Z*) tube lattice schematic. (**d**) Armchair (*A*) tube lattice schematic (along tube axis). (**e**) Schematics representing simulated inner and outer (DWCNT) end caps showing positions of 5-member rings (*pentagons*). These are the lowest energy DWCNT structural pairings. Note that designations are outer/inner tube convention rather than inner/outer tube convention discussed in text (Adapted from Lair et al. (2008))

Fig. 1d shows the corresponding armchair (A) configuration. Figure 1e shows end cap (nuclei) simulations for the lowest energy tubes for each structural type forming DWCNTs: Z/A (Z@A), A/A (A@A), Z/Z (Z@Z), and A/Z (A@Z), representing inner/outer end caps showing where the 5-member rings or pentagons are located

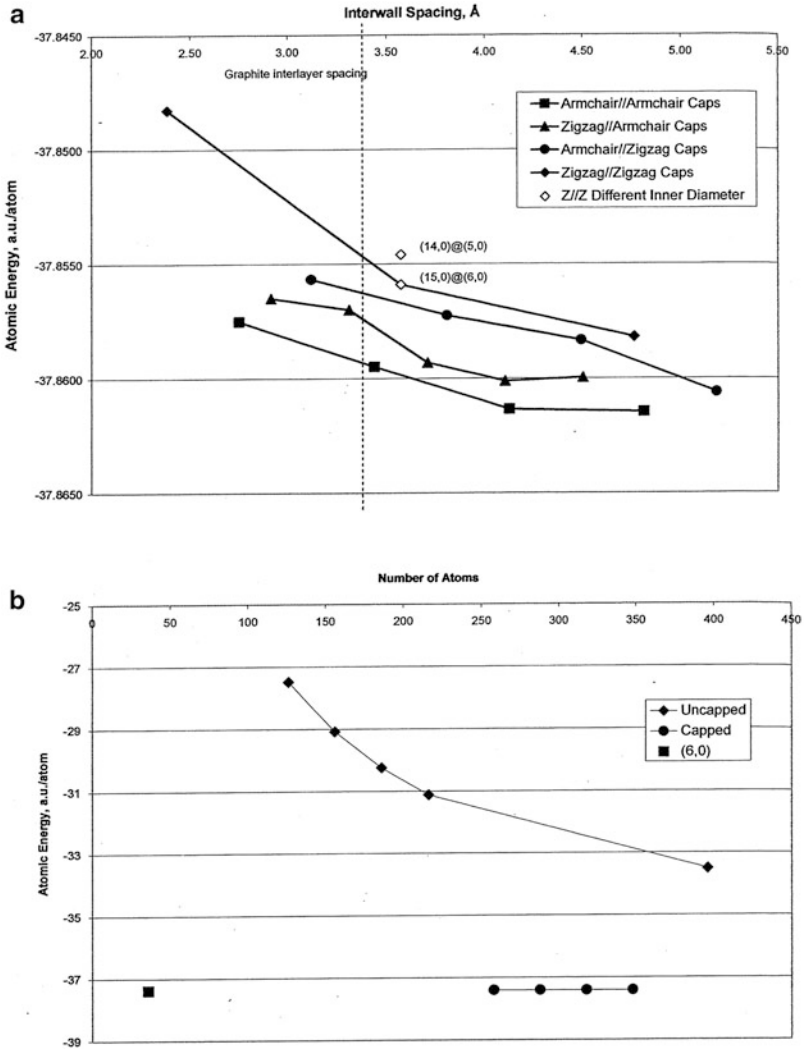


Fig. 2 (**a**) Atomic energy of capped DWCNTs versus interwall spacing (*IWS*). The graphite (or graphene) IWS shown in ~ 3.4 Å. (**b**) Atomic energy of capped and uncapped (6, o) @ (15, o): Z/Z DWCNTs versus outer tube length. Note that the designation in these figures represents outer tube/inner tube convention: Z/Z, outer @ inner (Adapted from Lair et al. (2008))

for each of the lowest energy DWCNTs for these structural pairings. In this configuration set (Fig. 1c), A@A: (5, 5) @ (10, 10) represents the lowest energy (Lair et al. 2008).

Figure 2a compares computed atomic energy for small DWCNTs having both ends capped, while Fig. 2b compares a least favorable Z/Z: (6, 0) @ (15, 0) DWCNT with ends uncapped (as in Fig. 1b) with both ends capped, as the tube

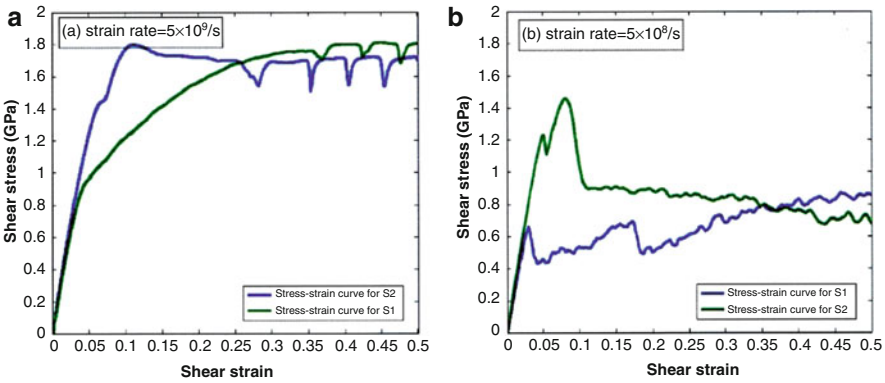


Fig. 3 Shear stress versus strain curves for nanolaminate S1 and S2 samples. (**a**) Shear strain rate of 5×10^9/s. (**b**) Shear strain rate of 5×10^8/s (Courtesy of Chen KG, Shi SQ, Zhu WJ, 2014, Interfacial behavior dominated plastic deformation in amorphous/crystalline nano laminates, unpublished, private communication)

length is increased, represented by an increasing, circumferential total number of carbon atoms. While for uncapped DWCNT growth the energy decreases with tube length (or number of carbon atoms), the growth for capped tubes is constant and considerably lower in energy than the uncapped tubes.

From the simulations and computations represented in ▶ Figs. 1e and ▶ 2 of chapter "Computer Simulation in Materials Science and Engineering," it is observed that carbon nanotubes nucleate and grow preferentially from caps (and cap types) and that A/A caps and corresponding CNT types are energetically favorable for DWCNT growth. Consequently, uncapped nanotube sections illustrated in ▶ Fig. 1b of chapter "Computer Simulation in Materials Science and Engineering" are untenable. It is also implicit from ▶ Fig. 2b of chapter "Computer Simulation in Materials Science and Engineering" that it is energetically favorable for long tubes to form since the energy decreases with tube length, and this may also apply to multiwall carbon nanotubes (MWCNTs).

While the atomistic computer simulations and related computations corresponding to Figs. 1 and 2 were characteristic of only several hundred atoms, molecular dynamics (MD) simulations of atomistic phenomena can involve hundreds of millions of atoms. Molecular dynamics simulations were utilized to investigate the properties of amorphous/crystalline interfaces (ACIs) and the effects and mechanical properties of nanolaminate ACIs. Two samples were constructed with the same sandwich geometry: 20 nm thick crystalline Cu sandwiched by 10 nm thick amorphous layers of $Cu_{40}Zr_{54}$, designated S1 and S2. Shear deformation, γ_{XZ}, was applied parallel to the ACI with periodic boundary conditions and different strain rates of 5×10^9/s and 5×10^8/s, respectively, as represented by the corresponding shear stress versus strain curves shown in Fig. 3a and b. To visualize the corresponding deformation map for the ACI under pure shear at these extreme strain rates, OVITO (Open Visualization Tool) MD software, a freely available visualization software

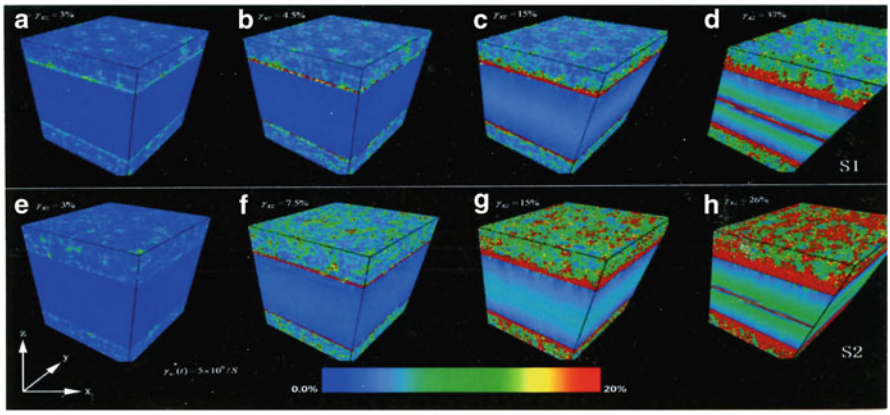


Fig. 4 Atomic shear strain maps for sample S1, (**a**) to (**d**), and S2, (**e**) to (**h**), for $\dot{\gamma}_{XZ}$ (t) = 5 × 10^9/s. OVITO MD simulations courtesy of K.G. Chen (Natl. Key Lab of shock wave and Detonation Physics, Mianyang, China; Dept. of Mech. Engr., Hong Kong Polytechnic Univ.), S.Q. Shi (Hong Kong Polytechnic), W.J. Zhu (Natl. Key Lab of Shock Wave and Detonation Physics) (Chen KG, Shi SQ, Zhu WJ, 2014, Interfacial behavior dominated plastic deformation in amorphous/crystalline nano laminates, unpublished, private communication). Note that strains in (**a**) to (**d**) correspond to 3 %, 4.5 %, 15 %, and 37 %, respectively. Strains in (**e**) to (**h**) correspond to 3 %, 7.5 %, 15 %, and 26 %, respectively, in upper left-hand corner

(Stukowski 2010; http://ovito.org) using c++ source code, was used to determine the atomic mechanisms responsible for the stress–strain curve features in Fig. 3. These simulations are shown in Figs. 4 and 5 corresponding to shear strain rates of 5 × 10^9/s and 5 × 10^8/s, respectively.

Shear transformation zones (STZs) appear in Figs. 4a, f and 5a, c with strains within the "elastic" regimes. In all cases the amorphous layers yield first before significant deviation from the linear portion of the stress–strain curves. The early presence of STZs demonstrates that the ACIs generate heterogeneity in the amorphous layers; however, the absence of shear bands at these extreme shear strain rates only causes small deviations from linear behavior in the shear stress–strain curves in Fig. 3. The first significant stress–strain transition points for S1 and S2 at γ_{XZ} (t) = 5 × 10^9/s are γ_{XZ} = 4.15 % and γ_{XZ} = 6.35 %, respectively. Interfacial sliding is responsible for this transition, and microbands (red atoms in Figs. 4 and 5) with highly localized atomic shear strain are induced. Note that S1 has much lower interfacial strength than S2 because of the excess free volume at the ACI as a consequence of insufficient quenching, and the glass–glass interfaces in the amorphous layer of S1 provide more STZ formation sites than in S2, making it more ductile. Following the microband formation, the rate of thickening is much smaller for shear deformation at 5 × 10^9/s than for 5 × 10^8/s.

In conclusion, Figs. 3, 4, and 5 illustrate that interfacial behavior, including interfacial sliding and microband thickening with highly localized shear strain,

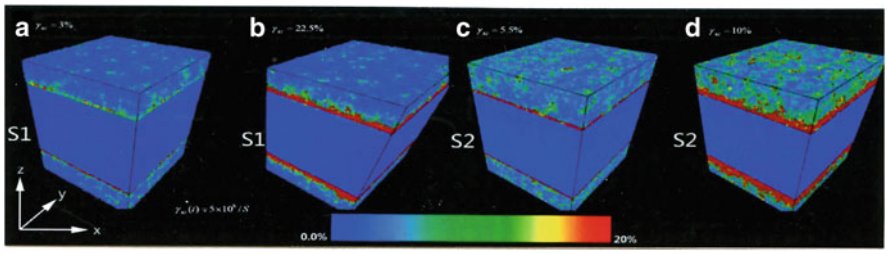


Fig. 5 Atomic shear strain maps for sample S1, (**a**) and (**b**), and S2, (**c**) and (**d**), for $\dot{\gamma}_{XZ}$ (t) = 5 × 10^8/s (Adapted courtesy of Chen KG, Shi SQ, Zhu WJ, 2014, Interfacial behavior dominated plastic deformation in amorphous/crystalline nano laminates, unpublished, private communication). Note that strains in (**a**) and (**b**) correspond to 3 % and 22.5 % (S1) while (**c**) and (**d**) correspond to 5.5 % and 10 %, respectively, in upper left-hand corner

dominates plastic deformation of an amorphous/crystalline layer under pure shear deformation. The corresponding simulations visualized in Figs. 4 and 5 provide a new multi-yielding scenario for such composites.

It can be appreciated that there is a considerable length-scale contrast between the atomistic views of a shear band visualized in Figs. 4 and 5 in contrast to the microscopic views of shear bands shown in ▶ Fig. 39 in chapter "Ballistic and Hypervelocity Impact and Penetration." This contrast will be even more obvious when considering, retrospectively, the other examples provided in this chapter.

Zhang et al. (2012) have recently used atomistic simulations using the embedded atom method (EAM) to study the nucleation of lattice dislocations from non-isostructural phase (Ca–Nb) interfaces with a *Kurdjumov–Sachs (K–S) orientation relationship*. The simulations revealed the role of atomic interface structure in determining the preferred nucleation sites and slip systems. Similarly, Tolvanen and Albe (2013) have used molecular dynamics (MD) simulations involving a modified version of the LAMMPS simulation package (Plimpton 1995) to model interatomic interactions of Cu atoms modeled by an EAM-type interatomic potential developed by Mishin et al. (2001); Fig. 6 illustrates the features of this simulation involving the plasticity of Cu nanocrystallites in nanoextrusion. Depending on the actual radius of the extrusion orifice, the mode of plasticity is either displacive or driven by the so-called surface amorphization. Figure 6a–d illustrates the maximum shear component in the initial system at the onset of plasticity for different orientations for a 15.6 Å orifice. Figure 6e–h shows the nucleation of the first partial dislocation which is met by a third partial nucleated at the opposite side of the orifice forming a <011>/6 stair-rod dislocation. In the case of the extremely stressed nanoparticle, the stair-rod dislocations do not block plastic flow as in bulk Cu because the continuous nucleation interactions make the Shockley partial-stair-rod systems unstable. The reader is referred to the original article by Tolvanen and Albe (2013) for additional details. In addition, many atomistic simulation images may be found on Google Images.

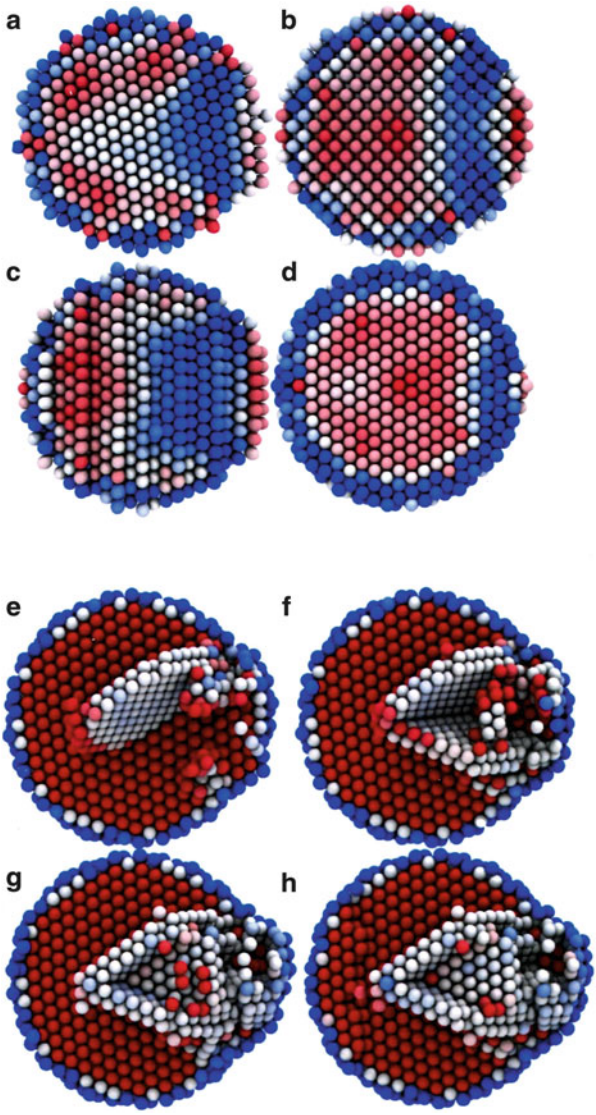


Fig. 6 Cu nanoparticle extrusion plasticity simulations. (**a**) to (**d**) show maximum shear component of atomic stress tensor expressed by coloring of atoms (*red* = 0 to > 1 GPa (*blue*)): (**a**) to (**c**) 15.6 Å orifice; (**d**) 11.6 Å orifice. (111) Planes are oriented parallel to extrusion direction in (**a**), tilted in (**b**), and perpendicular in (**c**). (**e**) to (**h**) show extrusion from the 15.6 Å orifice: (**e**) shows first Shockley partial nucleation; (**f**) two non-locking Shockley partials; (**g**) formation of stair-rod dislocation; (**h**) dislocations are pinned to the particle surface through a <312 > dislocation (Adapted from Tolvanen and Albe (2013))

Microstructure Simulations

Just as there are currently a great number of atomistic simulations such as those illustrated in Figs. 4, 5, and 6, there are also great numbers of larger-scale simulations of microstructures and associated phenomena, including dynamic events employing video simulations and other microstructure manipulations involving

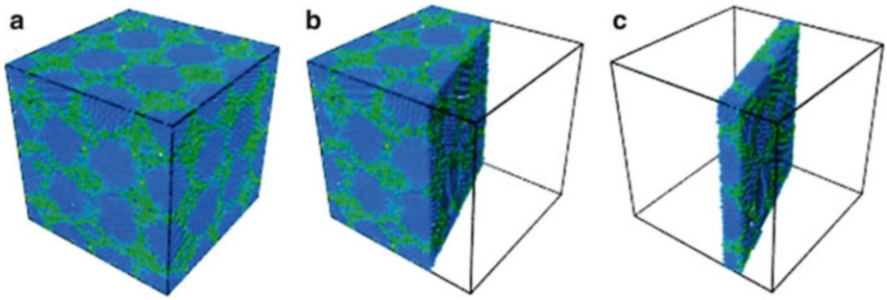


Fig. 7 MD (OVITO) simulation of polycrystalline volume sections (see Stukowski 2010) (Adapted from (http://www.ovito.org/manual/particles.modifiers.slice.html)). (**a**) Original volume. (**b**) Cut volume. (**c**) Sliced volume. Slicing can involve any orientation

3D-materials science and engineering concepts. Figure 7 illustrates these concepts very generally. Beginning with Fig. 7a showing an atomistic 3D section simulating a polycrystalline grain structure (which can be scaled from the nano-regime (<10 nm grain size) to the micro-regime (>10 μm grain size)), the section can be manipulated by deleting structure from the one side (Fig. 7b) or cutting a slice as illustrated in Fig. 7c. R.F. Zhang and T.C. Germman of Los Alamos National Laboratory have posted an MD simulation movie of shock loading of a polycrystalline section similar to Fig. 7a (see http://zrflanl.files.wordpress.com/2012/02/spasmfigmodel.jpg).

Microstructure–Mesostructure Simulation Features

In connection with atomistic simulations in Figs. 6 and 7, microstructure or scale enlargements can be extended to the tens of microns, while more practical engineering simulations make it difficult to visualize atomistic features in this context. In addition, atomistic simulation codes or software are incapable of visualization at the mesoscale. As noted in chapter "► Computer Simulation in Materials Science and Engineering," even mesoscale simulations involving FE software require mesh selections appropriate to visualization at specific length scales. In some cases, microstructures can be assumed and validated by direct observations utilizing appropriate length-scale resolution, e.g., optical microscopy for micron-scale microstructures versus scanning or transmission electron microscopy for submicron or nanoscale structures.

Figure 8 presents an example of this phenomenon. Figure 8a shows a half section of a copper plate (ballistic target plate) penetrated by a tungsten heavy alloy (WHA: 93 % W, 4.9 % Fe, 2.1 % Ni) rod projectile (7.8 cm in length and 0.78 cm in diameter: d_p) impacting the target plate at 1.5 km/s. Debris and surface structure or microstructure on the penetration channel wall is shown by dotted circles in Fig. 8a, while Fig. 8b shows an optical microscope image of these surface features composed of WHA debris fragments eroded from the WGA projectile during

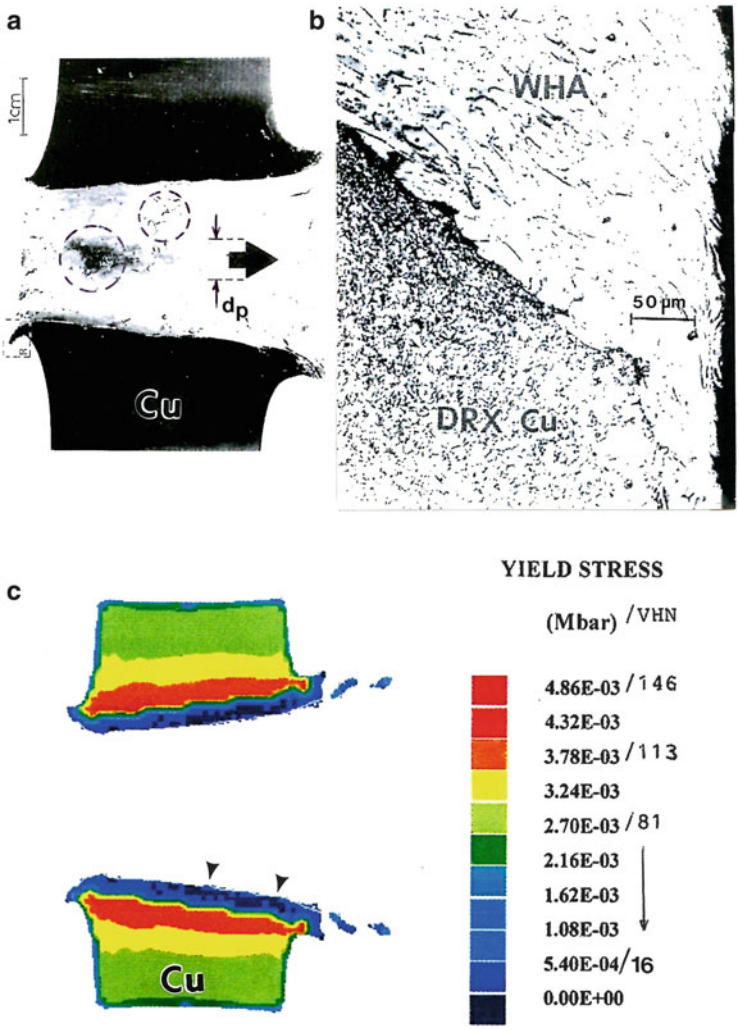


Fig. 8 Comparison of experimental and computer simulated impact and penetration phenomena. (**a**) Half-section of Cu target plate penetrated by WHA rod projectile at 1.5 km/s. The rod diameter (d_p) is shown scaled by *arrow*. (**b**) Optical microscope image typical for dotted circle zones in penetration wall of (**a**). (**c**) AUTODYN simulation of (**a**). *Arrows* show debris zones corresponding to *dotted circles* in (**a**). Note that color key shows residual yield stress which is converted to Vickers microindentation hardness values (VHN) by VHN $\cong 3\ \sigma_y$. 1 VHN = 0.01 GPa (Adapted from Kennedy and Murr (2002))

penetration, as well as a dynamic recrystallized (DRX) zone at the Cu target wall. AUTODYN-2D simulation of Fig. 8a in Fig. 8c shows the residual microindentation hardness values (VHN) in color zones corresponding to residual yield stress values in Mbar. The blue and very dark blue-to-black zone represents the soft, recrystallized (DRX) material at the penetration wall (arrows in Fig. 8c) as

well as debris (DRX Cu and WHA material) at the back surface or target exit surface (see ▶ Figs. 28 and ▶ 29 of chapter "Ballistic and Hypervelocity Impact and Penetration"), corresponding to the microstructures shown in Fig. 8b.

The notable feature of the simulation in Fig. 8c is that the color zones exaggerate the true residual yield stress or microindentation hardness regimes described in more detail in related examples discussed and illustrated in chapter "▶ Ballistic and Hypervelocity Impact and Penetration." Nonetheless, the implications for microstructural regimes are implicit in the simulation color renderings, although not to scale.

Simulations of thermodynamic phenomena across the length scales pose another important application in materials science and engineering. Such phenomena include calculated or simulated phase diagrams, especially fundamental binary as well as ternary diagrams and solidification phenomena which span the atomic to the extreme macroscale: lengths in tens of meters and corresponding weights in tens of tons. Computational thermodynamics based on calculation of phase diagram (CALPHAD) approaches developed in the last decades have enabled materials scientists and engineers to make routine phase stability calculations and phase equilibria simulations (Kaufman and Bernstein 1970; Wang et al. 2013). In such calculations and simulations, verification and validation can be achieved by comparison with the wide range of experimentally determined phase diagrams, both binary and tertiary (chapter "▶ Computer Simulation in Materials Science and Engineering").

A contemporary, integrated software package for multicomponent phase diagram calculation called PANDAT™ bridges the thermodynamic calculation, property optimization, and kinetic simulation of multicomponent systems based on CALPHAD (Chen et al. 2002; Cao et al 2009). This software package, combined with thermodynamic/kinetic/thermophysical databases, provides an integrated format for phase diagram calculation and materials property simulations of multicomponent systems for materials selection and design as well as product performance and failure analysis. As implicit in considering ▶ Eqs. 14, 15, 16, and 17 in chapter "Computer Simulation in Materials Science and Engineering," the CALPHAD approach requires fundamental information of the Gibbs free energies of all phases as a function of temperature and chemical composition. Simulations at temperature, especially ternary phase diagram calculation, result in phase-field structures corresponding to variations in temperature.

As illustrated, for example, in chapter "▶ Examples of Directional Crystal Structures: Gas Turbine Component Applications in Superalloys" (Figs. 5, 7, and 8), nickel-based superalloys have dominated high-temperature aircraft and aerospace applications involving turbine blades and related components for decades. However, higher performance systems require operation at higher temperatures ($>$1,150 °C). Next-generation jet engines, for example, will require alloys capable of sustained operation at 1,300 °C. Niobium and molybdenum-based alloys have been explored because they offer higher melting points and extended creep resistance when alloyed with Si and other oxidation-resistant elements such as B, Cr, Ti, and Al. The use of computed (simulated) isothermal sections or phase diagrams has facilitated the exploration of such higher-temperature alloy systems (Ventura et al. 2009; Natividad et al. 2010; Portillo and Varma 2012; Dasary and Varma 2013).

An example of the use of PANDAT software (CompuTherm LLC) in materials design and development is illustrated in Figs. 9 and 12. Figure 9 and 10 show calculated isothermal sections (ternary phase diagrams) including phases or phase fields for Ni–Cr–Mo alloys with fixed additions of Re and Si (15 % Re, 15 % Si) at various temperatures. Correspondingly, Figs. 11, 12, and 13 show the microstructures and phase compositions observed by electron backscatter imaging in the SEM. Predicted microstructural components at 25 °C in Fig. 9a include Cr NbSi, sigma phase, α-solid solution, and Nb_5Si_3. As shown in Fig. 11a, the as-cast microstructure exhibits the presence of $NbCr_2$ (Laves phase), α-(Nb) solid solution (Nb_{ss}), Re_2Si, and $(Nb, Mo)_5 Si_3$. The Re_2Si is formed around the solid solution and would be expected to prevent oxygen ingress (oxidation).

Figures 9b and 10 illustrate the variation of phase fields with temperature, while the composition within the shaded region remains fixed. No phase transformations are indicated in Figs. 9 and 10.

Figure 11b illustrates a decrease in the silicide phase size and an increase in the Re_2Si surrounding the Nb solid solution, while at 1,000 °C (Fig. 12a) a eutectic-like microstructure appears along with $NbCr_2$. At 1,400 °C Re_2Si and $NbCr_2$ are noticeably increased in the Re_2Si layer surrounding the Nb solid solution phase.

At the exterior of the oxidizing alloy, $CrNbO_4$ begins to form at 900 °C which retards oxidation to some extent along with Nb_5Si_3 which is promoted by Mo addition (Dasary and Varma 2013). Other oxides formed include Nb_2O_5, MoO_3, and SiO_2.

Macroengineering and Macrosimulation

Although solidification is a very complex process (Kurz and Fisher 1989), the use of computational methods to investigate both theoretical and practical aspects of solidification has been very successful for decades in understanding microstructures as these relate to materials properties and performance, especially in the context of process (solidification) phenomena. To understand solidification fundamentals, two modeling techniques have emerged: molecular dynamics (MD) and phase-field (PF) modeling for complex geometries and topology, such as dendrites (Humadi et al. 2013). Various modeling techniques at the macroscale based on FE and computational fluid dynamics have become reliable tools to investigate die filling and subsequent solidification processes in components and products ranging from automotive components to aeroengine turbine blades described in chapters "► Examples of Directional Crystal Structures: Gas Turbine Component Applications in Superalloys" and "► Chemical and Electrochemical Degradation and Failure of Materials," for example.

Li et al. (2013) have recently reviewed process modeling and simulations of heavy castings and forgings. One of the notable examples described involves simulation of Francis hydraulic turbine runner blades for the Three Gorges Power Station in the Yangtze Gorges in the upper reaches of the Yangtze River forming a common boundary for Hubei Province, Sichuan Province, and Chongqing City in

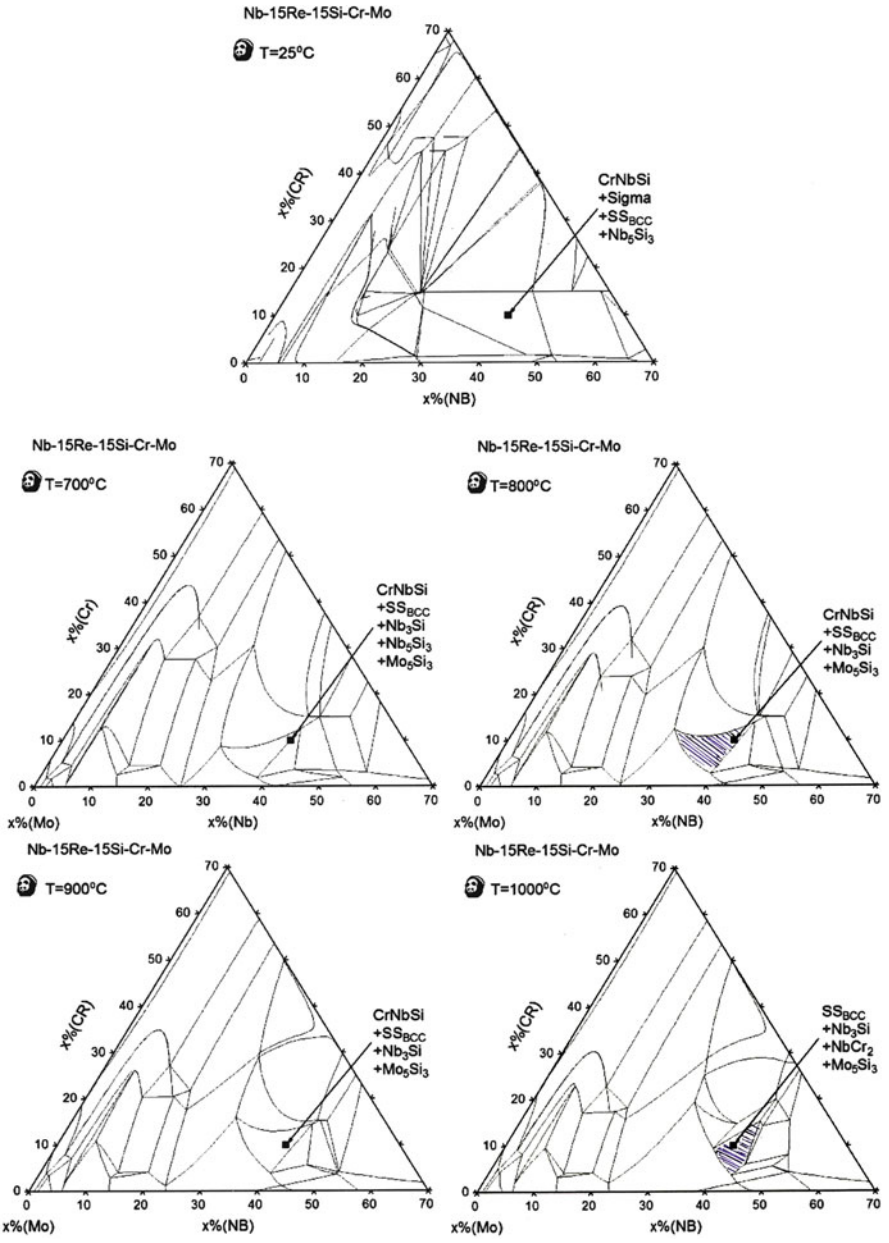


Fig. 9 Isothermal sections for Nb–Re–Si–Cr–Mo from 25 °C to 1,000 °C; computed using PANDAT software (section at 25 °C is adapted from Dasary and Varma (2013). Remaining sections are courtesy of R.M. Dasary and S.K. Varma, UTEP)

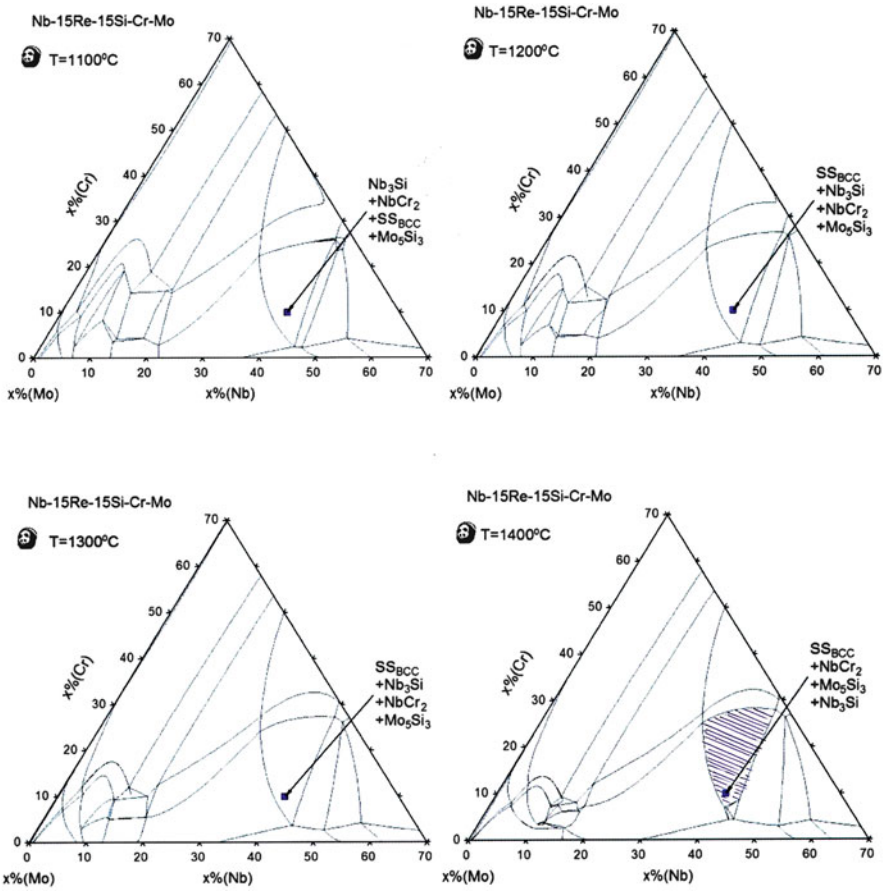


Fig. 10 Isothermal sections for Nb–Re–Si–Cr–Mo from 1,100 °C to 1,400 °C; computed using PANDAT software (Courtesy of R.M. Dasary and S.K. Varma, UTEP)

China. Figure 13a illustrates the 2.4 km hydropower generation dam, while Fig. 13b shows one of 32 runners being inserted into a hydroturbine where water flow is converted into mechanical and correspondingly electrical energy. These runners were 10.6 m in diameter and 5 m high, weighing 425 t. Each runner as shown in Fig. 13b consists of a top crown and a lower band through the blades which are welded together. When fully operational, the project produces 22.5 GW of electricity, nearly 10 % of China's total installed hydroelectric capacity of 230 GW, which accounts for ~ 20 % of China's overall power generation mix.

The processing of individual blades included casting, shakeout, cutting off the gating system, and subsequent heat treatment. Consequently, blade deformation was severe and unpredictable during hot processing, and blades were manufactured to allow for large machining to assure proper dimensions. Cast and annealed (heat-

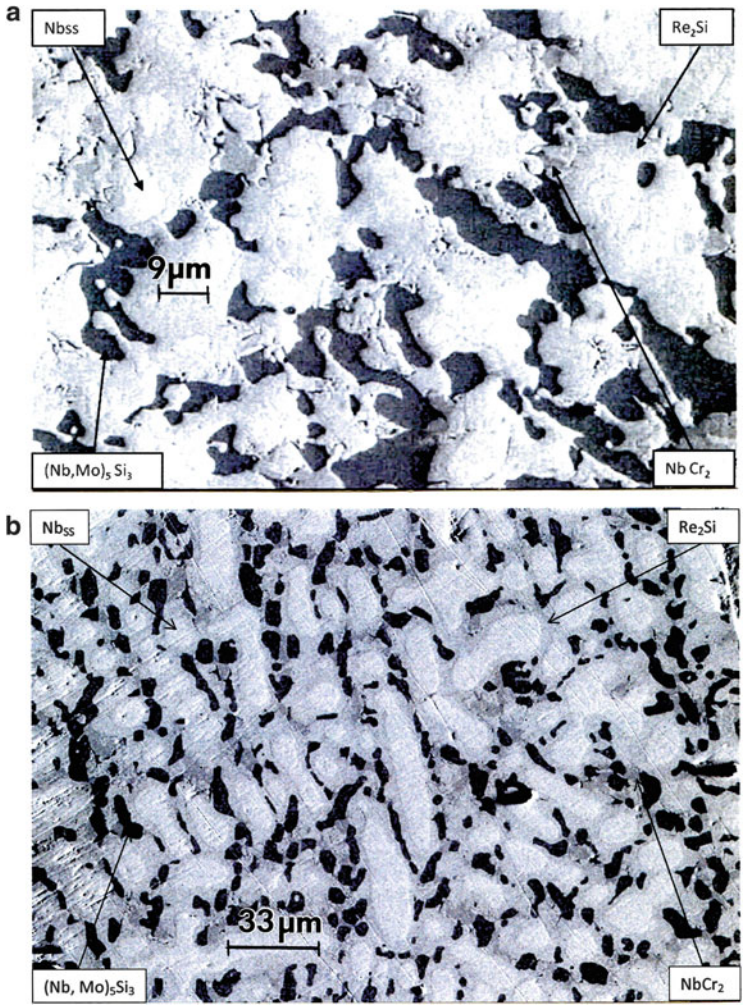


Fig. 11 SEM (electron backscatter) views of Nb–15Re–15Si–10Cr–20Mo alloy microstructures. (**a**) As cast (Adapted from Dasary and Varma (2013)). (**b**) 800 °C (Courtesy of R.M. Dasary and S.K. Varma, UTEP)

treated) stainless steel blades (18.3 Cr, 4 Ni, 1.7 Cu, 1.8 Mn balance, Fe; 40 % martensite, 35 % ferrite, 35 % austenite phase mixture nominally) were welded to the cast ring and crown.

In order to predict the deformation of the blade accurately and give a reasonable machining allowance, an integral finite element method model based on ProCast software was developed to simulate the deformation of the blade during the entire hot procedure including casting, shakeout, cutting off the gating system, and heat treatment. Two important things were considered in this simulation model. The first one is that some of the thermophysical parameters, such as expansion coefficient,

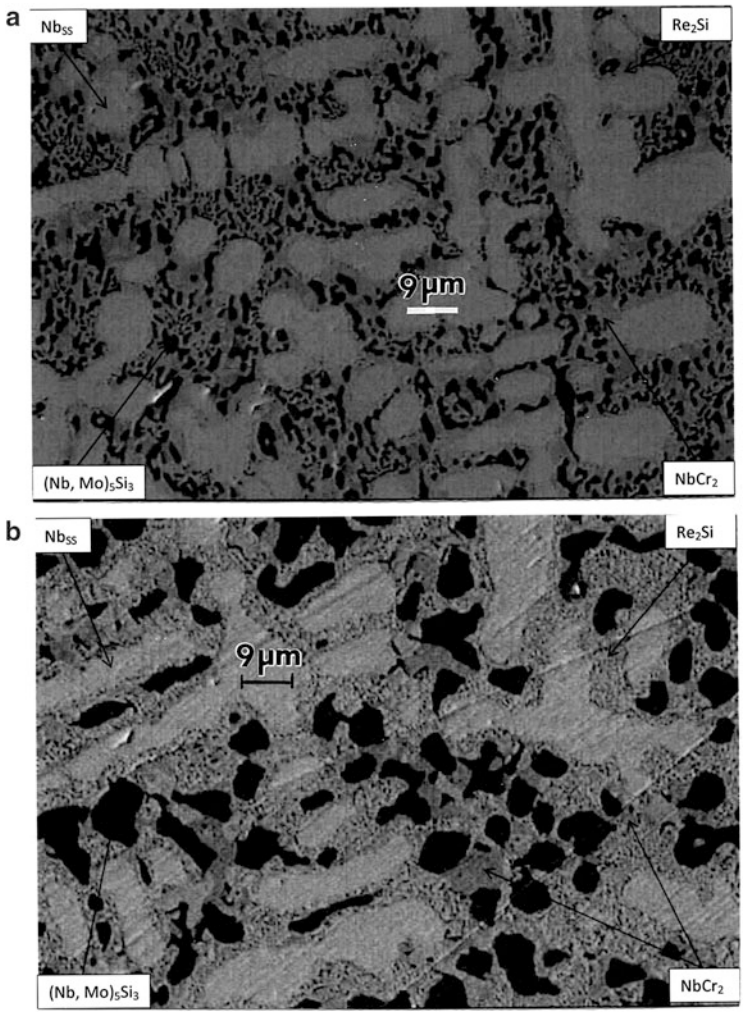


Fig. 12 SEM (electron backscatter) views of Nb–15Re–15Si–10Cr–20Mo alloy microstructures. (**a**) 1,000 °C. (**b**) 1,400 °C (Courtesy of R.M. Dasary and S.K. Varma, UTEP)

Young's modulus, and strength of the material at the same temperature, are different between heating and cooling processing during the heat treatment because the microstructure of the material is different during heating and cooling. The second is that the stress and strain in the preceding procedure should be accurately transferred to the following step, especially when some components, i.e., mould, gating system, and riser, are removed or deleted. In this simulation model, the interaction between the removed component and remaining component is equivalent to an external force when the component is deleted from the system.

Fig. 13 Three Gorges Dam Project, Yangtze River, China (**a**) (Photo courtesy of nipic.com). (**b**) shows a 425 t Francis turbine runner being inserted into the turbine housing (Courtesy of Alstom Turbine Technology Center, France; the manufacturer)

Figure 14 shows the displacements of the blade after solidification and cooling in the sand mould. The displacements can be divided into two different kinds. The deformation along x- and y-axes is mainly contributed by the shrinkage; therefore, this deformation can be counteracted by adding a proper shrinking gauge on the wood pattern. However, the distorted deformation along the z-axis should be treated deliberately. A proper reversed deformation should be considered. Figure 15 shows the displacements of the blade after heat treatment. The deformation tendency of the blade in Figs. 14 and 15 is almost the same; however, the numerical values and distribution of the displacement change remarkably.

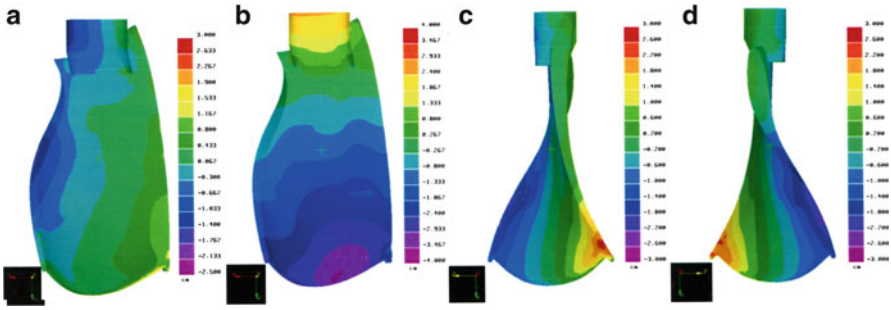


Fig. 14 Turbine runner blade displacements after solidification and cooling in a sand mould along the x-axis (**a**), y-axis (**b**), along the z-axis at the crown edge (**c**), and the band edge (**d**) (Adapted from Wang et al. (2010). Courtesy of the authors)

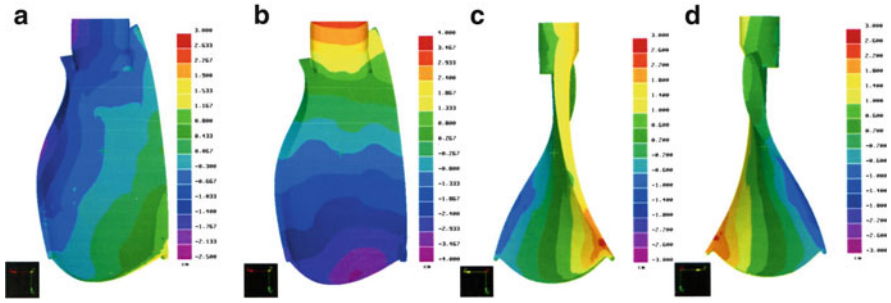


Fig. 15 Turbine runner blade displacements after heat treatment along the x-axis (**a**), the y-axis (**b**), the z-axis at the crown (**c**), and at the band edge (**d**); corresponding to cast blade in Fig. 14 (Adapted from Wang et al. (2010). Courtesy of the authors)

Closure

In retrospect, a few examples of computational materials science and engineering along with computer simulations have been presented in this chapter which represent the length scales and simulation methods illustrated in ▶ Fig. 2 in chapter "Computer Simulation in Materials Science and Engineering." There are of course many thousands of examples of length-scale simulations as well as timescale simulations (including video simulations) accessible on the Internet.

While this chapter concludes this handbook, it can, in retrospect, provide a concluding glimpse of the vastness of materials and materials systems as these relate to their structures, properties, processing, and performance issues characteristic of materials science and engineering.

References

Cao W, Chen S-L, Zhang F, Wu K, Yang Y, Chang YA, Schmid-Fetzer R, Oates WA (2009) PANDAT software with PanEngine, PanOptimizer and PanPrecipitation for multi-component phase diagram calculation and materials property simulation. Calphad Comp Coupling Phase Diagr Thermochem 33:328–342

Chen SL, Daniel S, Zhang F, Chang YA, Yan X-Y, Xie F-Y, Schmid-Fetzer R, Oates WA (2002) The PANDAT software package and its applications. Calphad 26(2):175–188

Dasary RM, Varma SK (2013) Short-term oxidation response of Nb-15Re-15Si-10Cr-20Mo alloy. J Mater Res Technol 3:25–34

Humadi H, Ofori-Opuka N, Provatas N, Hoyt JJ (2013) Atomistic modeling of solidification phenomena using the phase-field crystal model. JOM 65(9):1103–1110

Kaufman L, Bernstein H (1970) Computer calculations of phase diagrams. Academic, New York

Kennedy C, Murr LE (2002) Comparison of tungsten heavy-alloy rod penetration into ductile and hard metal targets: Microstructural analysis and computer simulations. Mater Sci Eng A325:131–143

Kurz W, Fisher DJ (1989) Fundamentals of solidification. Trans Tech Publications, Zurich

Lair SL, Herndon WC, Murr LE, Quinones SA (2006) End cap nucleation of carbon nanotubes. Carbon 44(3):447–455

Lair SL, Herndon WC, Murr LE (2008) Stability comparison of simulated double-walled carbon nanotube structures. Carbon 46:2083–2095

Li D, Sun M, Wang P, Kang X, Fu P, Li Y (2013) Process modelings and simulations of heavy castings and forgings. AIP Conf Proc 1532(1):81–94

Mishin Y, Mehl MJ, Papaconstantopoulos DA, Voter AF (2001) Structural stability and lattice defects in copper: ab initio, tight-binding, and embedded atom calculations. Phys Rev B 63:224106–1, 1–6

Natividad S, Acosta A, Amato K, Ventura J, Portillo B, Varma SK (2010) Heat treatment and oxidation characteristics of Nb-20Mo-15Si-5B-20 (Cr, Ti) alloys from 700 to 1400 °C. Mater Sci Forum 638–642:2351–2356

Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117(1):1–19

Portillo B, Varma SK (2012) Oxidation behavior of Nb-20Mo-15Si-25Cr and Nb-20Mo-15Si-25Cr-5B alloys. Metall Mater Trans A 43A:147–154

Stukowski A (2010) Visualization and analysis of atomic simulation data with OVITO – the open visualization tool. Model Sim Mater Sci Engr 18:015012–015020

Tolvanen A, Albe K (2013) Plasticity of Cu nanoparticles: dislocation-dendrite-induced strain hardening and a limit for displacive plasticity. Beilstein J Nanotechnol 4:173–179

Ventura J, Portillo B, Varma SK (2009) The oxidation resistance of Nb-20Mo-15Si-5B-20Cr up to 1300 °C. J Metall 7:72–75

Wang P, Xiao N, Li D, Li Y (2010) Numerical simulation of deformation during hot procedure for large hydraulic turbine runner blade. Mater Sci Forum 654–656:1565–1569

Wang Y, Shang S, Chen L-Q, Liu Z-K (2013) Density functional theory-based database development and CALPHAD automation. JOM 65(11):1533–1539

Zhang RF, Wang J, Beyerlein IJ, Misra A, German TC (2012) Atomic-scale study of nucleation of dislocations from fcc-bcc interfaces. Acta Mater 60:2855–2865

Index

L.E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*,
DOI 10.1007/978-3-319-01815-7